Stable mixture GARCH models

Broda, Simon A ; Haas, Markus ; Krause, Jochen ; Paolella, Marc S ; Steude, Sven C

Abstract: A new model class for univariate asset returns is proposed which involves the use of mixtures of stable Paretian distributions, and readily lends itself to use in a multivariate context for portfolio selection. The model nests numerous ones currently in use, and is shown to outperform all its special cases. In particular, an extensive out-of-sample risk forecasting exercise for seven major FX and equity indices confirms the superiority of the general model compared to its special cases and other competitors. Estimation issues related to problems associated with mixture models are discussed, and a new, general, method is proposed to successfully circumvent these. The model is straightforwardly extended to the multivariate setting by using an independent component analysis framework. The tractability of the relevant characteristic function then facilitates portfolio optimization using expected shortfall as the downside risk measure.

DOI: https://doi.org/10.1016/j.jeconom.2012.08.012

Originally published at:
Broda, Simon A; Haas, Markus; Krause, Jochen; Paolella, Marc S; Steude, Sven C (2013). Stable mixture GARCH models. Journal of Econometrics, 172(2):292-306.
DOI: https://doi.org/10.1016/j.jeconom.2012.08.012

Stable Mixture GARCH Models*

Simon A. Broda[a] Markus Haas[b]

Jochen Krause[c] Marc S. Paolella[c,d,†] Sven C. Steude[c]

[a] *Department of Quantitative Economics, University of Amsterdam, The Netherlands*

[b] *Institute for Quantitative Business and Economics Research, University of Kiel, Germany*

[c] *Department of Banking and Finance, University of Zurich, Switzerland*

[d] *Swiss Finance Institute*

September 4, 2012

Abstract

A new model class for univariate asset returns is proposed which involves the use of mixtures of stable Paretian distributions, and readily lends itself to use in a multivariate context for portfolio selection. The model nests numerous ones currently in use, and is shown to outperform all its special cases. In particular, an extensive out-of-sample risk forecasting exercise for seven major FX and equity indices confirms the superiority of the general model compared to its special cases and other competitors. An improved method (in terms of speed and accuracy) is developed for the computation of the stable Paretian density. Estimation issues related to problems associated with mixture models are discussed, and a new, general, method is proposed to successfully circumvent these. The model is straightforwardly extended to the multivariate setting by using an independent component analysis framework. The tractability of the relevant characteristic function then facilitates portfolio optimization using expected shortfall as the downside risk measure.

Keywords — Density Forecasting; Expected Shortfall; Fat Tails; ICA; GARCH; Mixtures; Portfolio Selection; Stable Paretian Distribution; Value-at-Risk.

JEL Classification: C13, C16, C22, C32, G17.

---

*Part of the research of Paolella has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. The research of Haas was supported by the German Research Foundation (DFG).

The authors are grateful to Kerstin Kehrle, Mico Loretan, Benoît Mandelbrot, J. Huston McCulloch, John P. Nolan, Eric Renault, Gennady Samorodnitsky, Murad S. Taqqu, David Veredas, Casper G. de Vries, and two anonymous referees, for providing an array of insightful and detailed comments and suggestions which led to a greatly improved paper.

†Corresponding author: marc.paolella@bf.uzh.ch

1 Introduction

Starting with the pioneering works of Mandelbrot (1963) and Fama (1965), a variety of studies have investigated the use of the stable Paretian distribution for modeling the unconditional distribution of asset returns. Given the very fat-tailed nature of weekly, daily, and higher-frequency financial returns data, it is not surprising that the stable distribution has been very successful in this regard. While there exist other fat-tailed, asymmetric distributions which also fit asset returns well, (see, e.g., Knight and Satchell, 2001, Paolella, 2007, and the references therein), none of these are closed under summation, a feature which is of great practical use in portfolio allocation; see e.g., Doganoglu et al. (2007) and Giacometti et al. (2007).

It is well-known that the stable Paretian is the only valid distribution that arises as a limiting distribution of sums of independently, identically distributed (iid) random variables. Given that error terms in econometric models are usually interpreted as random variables that represent the sum of all the effects not being captured by the model, the use of the stable Paretian assumption should be highly desirable; for further discussion see McCulloch (1996), Rachev and Mittnik (2000), Rachev (2003) and Nolan (2012).

Despite the advantages of the stable distribution for modeling real phenomena, a popular, albeit often misguided, critique of the use of the stable Paretian distribution is the lack of existence of the second moment. (Recall that unless the tail index $\alpha \in (0, 2]$ equals two, the tails are so heavy that absolute moments of order α and higher do not exist.) A number of studies have attempted to measure the tail index of the distribution of financial returns, as information about the tail index can be used to derive the probability of large price movements and, especially, market crashes (Jansen and de Vries, 1991). Additionally, knowing the maximum existing moment of the return process is of interest, as the lack of second moments will have consequences for risk and portfolio analysis. Nevertheless, it has been demonstrated by several authors that this endeavor is extremely difficult, so that no conclusive evidence of whether second moments of daily financial returns exist or not has been presented. For example, Kratz and Resnick (1996) discuss the inevitable and potentially "outrageous" bias inherent in tail thickness estimators such as the (in)famous Hill estimator (Hill, 1975).

The extreme bias problem for the Hill estimator with stable Paretian distributions has been studied in depth by McCulloch (1997), Mittnik et al. (1998) and Weron (2001). These and related studies have caused a shift from tail estimation to the adoption of a fat-tailed parametric assumption, such as the stable distribution. Indeed, this is embodied in the bold statement of Adler (1997), who states that "*Many of the problems faced by the Hill and related estimators of the tail decay parameter α can be overcome if one is prepared to adopt a more parametric model and assume, for example, stable innovations*". He goes on to say that "*Overall, it seems that the time may have come to relegate Hill-like estimators to the Annals of Not-Terribly-Useful Ideas.*" With the above comments about tail estimation in mind (and Adler's positive stance towards use of the stable distribution), we know of no definitive way to test the existence or nonexistence of second moments. Based on the positive empirical findings below, we conclude that, for applications to density and risk forecasting, the question of existence of second moments is only of secondary interest. This stance is further supported by Malevergne et al. (2005) who state that "*for most practical applications, the relevant question is not to determine what is the true asymptotic tail, but what is the best effective description of the tails in the domain of useful applications*".

Another critique of the use of the stable Paretian distribution is the complexity of computing its density, as required for the likelihood—which for conditional (non-iid) models, is required for parameter estimation. With modern computing power, and the availability of several algorithms, this is no longer a hindrance. Nevertheless, there still appears to be no existing method which is both fast and delivers the high accuracy required for likelihood optimization. To this end, we propose yet another method for its computation. It capitalizes on the vectorized nature of modern computing languages and yields a method which is very fast, but also superior to existing algorithms in terms of accuracy. Its details, and references to other methods (and their flaws) are detailed below in Section 2.3.

The real problem with the use of the stable-Paretian, or *any* skewed, fat-tailed distribution for modeling the unconditional distribution of asset returns, is that they cannot capture the time-varying volatility so strongly evident in daily and higher-frequency returns data.[1,2] Section 2 discusses the use of the stable distribution in conjunction with GARCH models to overcome this limitation.

Another popular and successful approach to the unconditional modeling of asset returns and VaR prediction involves the use of finite mixtures of normal distributions. Owing to its great flexibility, a normal mixture, even with just two components, is well-suited for capturing the usual stylized facts typical in a financial context. This model has been motivated and investigated by numerous authors, including Kon (1984), who suggests that returns may be influenced by a series of different information flows including a non-information distribution, a firm-specific information distribution, and a market-wide information distribution—hence, a mixture of three normal distributions. A different economic motivation for the presence of a mixture of distributions is provided by Vigfusson (1997), who builds on theoretical work which explains the stylized facts of financial time series by the interaction of heterogeneous groups of agents, with the groups processing market information differently; see, e.g., Samanidou et al. (2007) for an overview of such models. This is in line with recent research with experimental data by Kirchler and Huber (2007), who show that heterogeneous fundamental information can be a major source for the emergence of fat tails and volatility clustering. The fact that, for each component, a central limit theorem argument can be used to justify the use of the normal distribution is appealing, and lends some theoretical justification for the model and its economic interpretations. The same holds for the use of stable distributions for the mixture components, via the generalized central limit theorem. Indeed, Salas-Gonzalez et al. (2009) propose mixtures of stable distributions with a view towards applications in engineering such as image and radar signal processing.

The aforementioned problem regarding stable distributions also applies to the use of mixtures: in an unconditional setting, the mixture cannot capture the strong time-varying volatility of the returns. This is addressed by several authors who combine mixture models with GARCH structures. In this paper, we propose a model which generalizes the normality assumption in the normal mixture GARCH model to allow for stable distributions, investigate some of its theoretical and empirical properties, and extend its use (in a limited way adequate for portfolio optimization) to a multivariate framework. Moreover, as the estimation of all mixture models,

---

[1] The term "volatility" is usually defined as the standard deviation of the log returns. In the models we shall be considering, the standard deviation is infinite, so this is strictly speaking a misnomer. In a slight abuse of terminology we shall continue to use the term and take it to refer to the (possibly time-varying) scale of the conditional return distribution.

[2] There is actually another problem with the stable Paretian (or mixtures thereof) which we do not address. The summability (or stability) property of the stable distribution and the definition of log returns implies that the tail index of the return distribution should remain the same at any frequency, i.e., intraday, daily, weekly, monthly, etc.. However, it is well-known that this is usually not the case, with, say, daily returns exhibiting a tail index considerably lower than two, but monthly data exhibiting nearly normal behavior. This occurs because, for such series, the returns are not iid stable Paretian, but rather have a distribution such that, via a central limit theorem, their sums approach normality. The iid aspect of this problem is addressed in Paolella (2001) by accounting for the non-constant scale term by application of a stable-GARCH filter and construction of a formal testing procedure, but even then, the null hypothesis of stability can be rejected for many (but not all) return series. This result is not in conflict with our stable mixture GARCH model because our goal is to (i) devise a model endowed with some plausible statistical and economic motivation, (ii) which yields relatively superior density and risk forecasts for daily (and possibly higher frequency) data, and (iii) can be used in a multivariate context via an ICA decomposition, but without concern for the stability (or lack thereof) aspect of returns. A possible way of incorporating all such features would be to use the tempered stable distribution, which also has a tractable characteristic function, mimics the shape of the stable distribution, but is such that, when iid copies are summed, the tail index increases; see Kim et al. (2008) and Kim et al. (2010).

and particularly mixture GARCH models, is numerically challenging due to the degeneracy problem, we introduce new estimators which elegantly resolve this. The degeneracy problem is illustrated graphically using real data, and the excellent performance of the new estimators is demonstrated via simulation.

The remainder of this paper is as follows. Section 2 introduces the new mixture GARCH model and the method we suggest for computing the stable density. The new estimators are presented in Section 3. Section 4 presents an empirical exercise. Section 5 details how the model lends itself to portfolio allocation. Section 6 concludes.

2 Stable Mixture GARCH

Among conditional volatility models, the normal-GARCH has proven itself to be highly effective, though despite its success in capturing a high percentage of the volatility movement, countless applications have confirmed that the residuals, or filtered innovations when applied to weekly, daily, or higher frequency asset return data, still deviate considerably from normality. This has given rise to a large number of alternative models which replace the normal distribution in the GARCH model by a fat-tailed, asymmetric one. Given its theoretical properties, the stable Paretian distribution suggests itself, as first proposed by McCulloch (1985). See Mittnik et al. (2002) for further references and technical details, and Mittnik and Paolella (2003) for a demonstration of its effectiveness in value at risk (hereafter VaR) forecasting.

2.1 Mixture GARCH

A model which addresses the fat-tailed, asymmetric innovation issue mentioned above, but also gives rise to rich volatility dynamics not possible in the traditional battery of GARCH models, involves the use of mixtures. Building on the success of the mixed normal distribution for capturing the unconditional skewness and excess kurtosis of asset returns, and on some special cases already in the literature, Haas et al. (2004b) and Alexander and Lazar (2006) independently propose a general model structure which endows each mixed normal component with a GARCH structure.

As in Haas et al. (2004b), we say that time series $\{\varepsilon_t\}$ is generated by a k-component mixed normal GARCH(r, s) process, denoted MixNormal-GARCH, if the conditional distribution of ε_t is a k-component mixed normal distribution with zero mean,

$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \text{MixNormal}\left(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t\right), \tag{1}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \ldots, \sigma_{k,t})'$ are column vectors, the mixed normal probability density function (pdf) is given by

$$f_{\varepsilon_t|\mathcal{F}_{t-1}}\left(x; \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t\right) = \sum_{i=1}^{k} \omega_i \phi\left(x; \mu_i, \sigma_{i,t}\right),$$

$\mathcal{F}_t$ represents the information available at date t, ϕ is the normal pdf, $\omega_i \in (0, 1)$ with $\sum_{i=1}^{k} \omega_i = 1$ and, to ensure zero mean, $\mu_k = -\sum_{i=1}^{k-1} \left(\omega_i/\omega_k\right) \mu_i$. The component variances $\sigma_{i,t}^2$ follow the GARCH-like structure

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\gamma}_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \tag{2}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,k})'$, $i = 0, \ldots, r$, are $k \times 1$ vectors, $\boldsymbol{\Psi}_j$, $j = 1, \ldots, s$, are $k \times k$ matrices, $\boldsymbol{\sigma}_t^{(\delta)}$ is short for $(\sigma_{1,t}^{\delta}, \sigma_{2,t}^{\delta}, \ldots, \sigma_{k,t}^{\delta})'$, and $\delta \in \mathbb{R}_{>0}$. The parameters of the model need to be such that $\boldsymbol{\sigma}_t^{(\delta)} > 0$, where, in case of non-scalars, $>$ indicates element-wise inequality. As discussed in the above references, we take $r = s = 1$ and restrict the $\boldsymbol{\Psi}_j$ to be diagonal, so in this case we will refer to the diagonal elements of $\mathbf{\Psi}_1$ as ψ_i, $i = 1, \ldots, k$. Further details, and a survey of the model and the extensions which followed are given in Haas and Paolella (2011).

Another feature of mixture GARCH models is their ability to contain conditional as well as unconditional component models. As it turns out, the component of the mixture assigned to the most volatile observations can often be adequately modeled by a component with a relatively high, but constant, variance—it does not require a GARCH structure. We identify such models, where only g, $g \leq k$, components follow a GARCH$(1,1)$ process, by appending the accessory (k, g). Thus, the k-component MixNormal-GARCH model given in (1), with $r = s = 1$ and $\mathbf{\Psi}_1 = \mathrm{diag}(\psi_1, \ldots, \psi_k)$, is referred to as MixNormal(k, g).

Similar to replacing the normal assumption in a standard GARCH model, it suggests itself to challenge the MixNormal assumption with alternative distributions in addition to, or instead of, increasing the number of mixture components, to achieve better out-of-sample forecasting performance. As demonstrated in Kuester et al. (2006), one or more of the component densities of the MixNormal model can still exhibit tails which are fatter than the normal. In Kuester et al. (2006) and Rombouts and Bouaddi (2009) the generalized exponential distribution (in short, GED) is applied and shown to lead to improved in-sample fits and quality of VaR forecasts. For comparison with the MixStable-GARCH model, and using the aforementioned (k, g) notation, we will refer to the MixGED-GARCH model as MixGED(k, g) in the following.

In our empirical section, we use the GED distribution given by the location-zero, scale-one pdf

$$f(x; p) = \frac{p}{2\Gamma(p^{-1})} \exp\left\{-|x|^p\right\}, \quad p > 0.$$

Its cumulative distribution function (cdf), required for VaR calculations, follows as

$$F(x; p) = \frac{1}{2}\left(1 - \bar{\Gamma}_{(-x)^p}\left(p^{-1}\right)\right), \quad x \leq 0,$$

where $\bar{\Gamma}$ is the incomplete gamma ratio, and, for $x > 0$, $F(x) = 1 - F(-x)$ due to symmetry.

2.2 The MixStable Model

Analogous to (1) and (2), we say that time series $\{\varepsilon_t\}$ follows a k-component mixed stable GARCH(r, s) process, denoted MixStable-GARCH, if the distribution of $\varepsilon_t \mid \mathcal{F}_{t-1}$ is a finite mixture of stable distributions. Its pdf is

$$f_{\varepsilon_t|\mathcal{F}_{t-1}}(x; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t) = \sum_{i=1}^{k} \omega_i f_S(x; \alpha_i, \beta_i, \mu_i, \sigma_{i,t}), \tag{3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$ is the set of tail indices, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is the set of asymmetry parameters corresponding to the k stable distributional components, and, as before, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)'$ is the set of weights, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ is the set of component location terms, $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \ldots, \sigma_{k,t})'$ is the set of strictly positive scale parameters, and $f_S(x; \alpha, \beta, \mu, \sigma)$ is the location-μ, scale-σ, stable Paretian pdf with tail index α and skewness parameter β, as in Samorodnitsky and Taqqu (1994). For the mixture GARCH, let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$ be the set of power GARCH coefficients. Further, we assume that $1 < \alpha_i \leq 2$, so that the mean exists, and, with $\alpha_{\min} = \min_i \alpha_i$, restrict $0 < \delta_i < \alpha_{\min}$, $i = 1, \ldots, k$. This is a natural extension of the power restriction in the stable-GARCH model as devised in Mittnik et al. (2002). If $\alpha_{\min} = 2$ (so that $\alpha_i = 2$ for all i), the δ_i need only be positive, $i = 1, \ldots, k$. As with MixNormal(k, g), to ensure zero mean, $\mu_k = -\sum_{i=1}^{k-1} (\omega_i/\omega_k)\, \mu_i$ is imposed. The component scale terms, analogous to the variance term in the MixNormal model, evolve according to

$$\boldsymbol{\sigma}_t^{(\boldsymbol{\delta})} = \boldsymbol{\gamma}_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \left|\varepsilon_{t-i}\right|^{(\boldsymbol{\delta})} + \sum_{j=1}^{s} \mathbf{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(\boldsymbol{\delta})}, \tag{4}$$

where $\boldsymbol{\sigma}_t^{(\boldsymbol{\delta})}$ is short for $(\sigma_{1,t}^{\delta_1}, \sigma_{2,t}^{\delta_2}, \ldots, \sigma_{k,t}^{\delta_k})'$. Motivated by our use of relatively (for GARCH applications) small sample sizes (see the comment in Footnote 11 and the discussion in Section 4.3 below), we impose $\alpha_i = \alpha_j$ and $\beta_i = \beta_j$ for all stable mixture components as well as $\delta_i = \delta_j$ for all mixture GARCH models in the following. As such, we drop the component index i and just write α, β and δ. Similar to MixNormal(k, g), MixStable(k, g) denotes the k-component mixed stable GARCH$(1, 1)$ process with diagonal $\boldsymbol{\Psi}_1$ matrix, and only g of the k components having a GARCH structure. For $\beta = 0$, the model nests the following:

1) The unconditional stable Paretian model, as proposed by Mandelbrot (1963) and Fama (1965), by taking $k = 1$ and no GARCH structure.

2) The unconditional mixed normal model from Fama (1965), Kon (1984) and others, by taking $\alpha_i =$ in each of the k components, and no GARCH structure.

3) The normal-GARCH model from Bollerslev (1986), by taking $k = 1$, $\alpha = 2$ and $\delta = 2$.

4) The symmetric stable-GARCH model of Mittnik et al. (2002), by taking $k = 1$.

5) The MixNormal(k, g) model (1) and (2) of Haas et al. (2004b) and Alexander and Lazar (2006), by taking $\delta = 2$ and $\alpha = 2$.

6) The "linear" two-component MixNormal-GARCH models of Vlaar and Palm (1993) and Bai et al. (2003).

The MixStable-GARCH model is very general; in particular, there are two sources of asymmetry in the model: μ_i and β. In the empirical section below, we will limit ourselves to two special cases, which we dub A^{1}MixStable(k, g) and A^{2}MixStable(k, g), respectively. In the former, we restrict β to zero, whereas in the latter, $\mu_i = 0$, so that only one source of asymmetry is considered at a time.

As the stable distribution does not possess a finite second moment if $\alpha < 2$, the MixStable-GARCH process will not be covariance stationary. It may still be strictly stationary, however. This follows from Liu (2007) who generalizes the Markov-switching GARCH(1,1) process of Haas et al. (2004a) to allow for more general power parameters and innovation distributions, and studies its dynamic properties. This process nests the model studied in this paper, and we state the stationarity condition for the parsimonious case where all the stable components are characterized by the same shape parameters α and β. It then follows from Corollary 2.1 of Liu (2007) that a sufficient condition for the process to be strictly stationary with a finite δth moment is that the eigenvalues of the matrix

$$\boldsymbol{\gamma}_1 \boldsymbol{\omega}' \kappa_{\delta,\alpha,\beta} + \boldsymbol{\Psi}_1 \tag{5}$$

are inside the unit circle, where, as in Mittnik et al. (2002),

$$\begin{aligned} \kappa_{\delta,\alpha,\beta} &= \eta_\delta^{-1} \Gamma\left(1 - \frac{\delta}{\alpha}\right) \left(1 + \tau_{\alpha,\beta}^2\right)^{\delta/(2\alpha)} \cos\left(\frac{\delta}{\alpha} \arctan\left(\tau_{\alpha,\beta}\right)\right), \\ \tau_{\alpha,\beta} &= \beta \tan\left(\alpha\pi/2\right), \\ \eta_\delta &= \begin{cases} \Gamma\left(1 - \delta\right) \cos\left(\frac{\pi\delta}{2}\right), & \text{if } \delta \neq 1, \\ \pi/2, & \text{if } \delta = 1. \end{cases} \end{aligned} \tag{6}$$

Term $\kappa_{\delta,\alpha,\beta}$ is the power-δ absolute moment of a stable random variable with tail index α and asymmetry parameter β; see Paolella (2007, Sec. 8.3) for a detailed derivation. For $\delta = 1$ and $\beta = 0$, (6) reduces to $\kappa_{1,\alpha,0} = 2\,\Gamma\left(1 - \alpha^{-1}\right)/\pi$.

Similar to the mixed normal GARCH model, condition (5) allows some (but not all) components to be driven by non-stationary GARCH dynamics, whereas overall the process will be stationary as long as the mixing weights of these components are sufficiently small. In particular, when $\max_j\{\psi_j\} < 1$ is satisfied, it follows from arguments similar to Haas et al. (2004a) that the eigenvalue condition is equivalent to

$$\sum_{j=1}^{k} \omega_j \frac{\kappa_{\delta,\alpha,\beta}\gamma_{1,j}}{1-\psi_j} < 1. \tag{7}$$

The ARCH(∞) representation of $\sigma_{j,t}^{\delta}$, given by

$$\sigma_{j,t}^{\delta} = \frac{\omega_j}{1-\psi_j} + \gamma_{1,j}\sum_{i=1}^{\infty}\psi_j^{i-1}|\varepsilon_{t-i}|^{\delta},$$

shows that the total impact of a shock on future volatility in regime j is

$$\gamma_{1,j}\sum_{i=0}^{\infty}\psi_j^{i} = \frac{\gamma_{1,j}}{1-\psi_j}.$$

Thus, condition (7) restricts the *average* total impact of a shock on the component-specific future volatilities.

2.3 Computational Remarks

From (3) and (4), it is readily apparent that the likelihood of the MixStable model is straightforward to calculate, provided a computable expression for the density of the stable Paretian distribution is available. Several authors have developed methods for this, including Doganoglu and Mittnik (1998), McCulloch (1998), Nolan (1998) and Mittnik et al. (1999). For evaluating (3), we compute the stable densities of the mixture components based on a variant of the real-valued integral expression of Zolotarev (1986) as given in Nolan (1997).[3]

2.3.1 Evaluation of the Stable Density

Considering existing first moments ($\alpha > 1$), the original expression of the stable pdf in Nolan (1997) reduces to

$$f_S(z,\alpha,\beta) = \frac{1}{\sigma}\begin{cases} \frac{\alpha}{\pi|\alpha-1|(z-\zeta)}\int_{-\tau}^{\pi/2} V(y;\alpha,\beta,z)\exp\{-V(y;\alpha,\beta,z)\}\,\mathrm{d}y, & \text{if } z>\zeta,\\ \Gamma(1+\frac{1}{\alpha})/\pi\cos(\tau)(1+\zeta^2)^{-1/(2\alpha)}, & \text{if } z=\zeta,\\ f_S(-z,\alpha,-\beta), & \text{if } z<\zeta, \end{cases} \tag{8}$$

where z is the standardized and transformed observation, $z = (x-\mu)/\sigma - \beta\tan(\alpha\pi/2)$, and

$$\begin{aligned} V(y;\alpha,\beta,z) &= \cos(\alpha\tau)^{1/(\alpha-1)}\left(\frac{(z-\zeta)\cos(y)}{\sin(\alpha y+\alpha\tau)}\right)^{\alpha/(\alpha-1)}\left(\frac{\cos(\alpha\tau+y(\alpha-1))}{\cos(y)}\right),\\ \zeta &= -\beta\tan\left(\frac{\pi\alpha}{2}\right),\\ \tau &= \frac{1}{\alpha}\arctan\left(\beta\tan\left(\frac{\pi\alpha}{2}\right)\right). \end{aligned}$$

For the fast computation of several evaluation points at once a vectorized implementation of (8) suggests itself, so we replace all mathematical operators by their element-wise counterparts. In doing so, we adapt a vectorized variant of the adaptive Simpson quadrature, (e.g., see Lyness, 1969) given by

$$\oint_{\boldsymbol{a}}^{\boldsymbol{b}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \begin{cases} (\boldsymbol{q}_1+\boldsymbol{q}_2)/2, & \text{if } \|\boldsymbol{q}_2-\boldsymbol{q}_1\|_\infty \le 10^{-6},\\ \oint_{\boldsymbol{a}}^{\boldsymbol{c}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \oint_{\boldsymbol{c}}^{\boldsymbol{b}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, & \text{otherwise,} \end{cases} \tag{9}$$

[3] We re-scale the distribution by $1/\sqrt{2}$, so that the standard normal distribution arises as a special case for $\alpha = 2$, or similarly for $p = 2$ in the GED case.

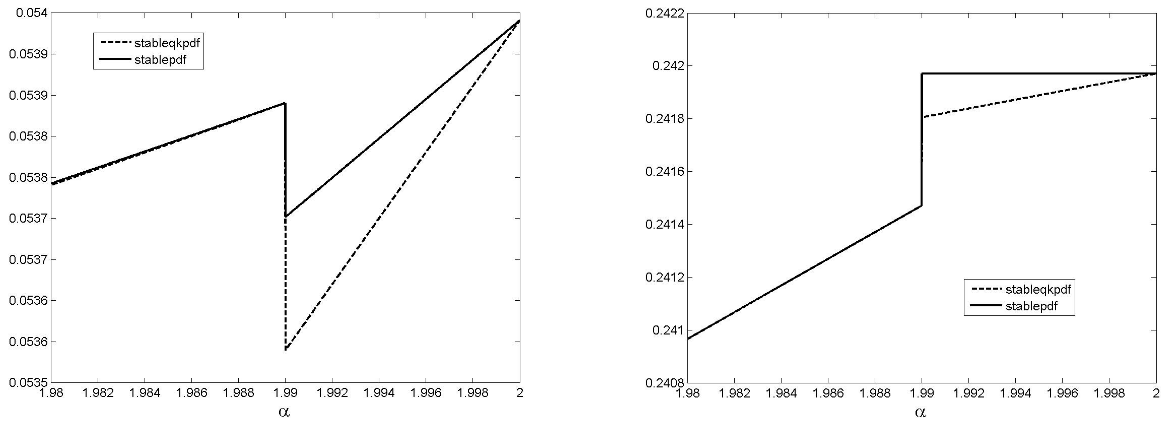
where $S\!\!\!\int$ denotes the (recursive) Simpson integral,

$$\boldsymbol{q}_1 = \frac{\boldsymbol{b}-\boldsymbol{a}}{6}\left(f(\boldsymbol{a}) + 4f(\boldsymbol{c}) + f(\boldsymbol{b})\right) \text{ and } \boldsymbol{q}_2 = \frac{\boldsymbol{b}-\boldsymbol{a}}{12}\left(f(\boldsymbol{a}) + 4f(\boldsymbol{d}) + 2f(\boldsymbol{c}) + 4f(\boldsymbol{e}) + f(\boldsymbol{b})\right),$$

as well as $\boldsymbol{c} = (\boldsymbol{a}+\boldsymbol{b})/2$, $\boldsymbol{d} = (\boldsymbol{a}+\boldsymbol{c})/2$ and $\boldsymbol{e} = (\boldsymbol{c}+\boldsymbol{b})/2$. Besides the reduced computational overhead by virtue of vectorization, the computation is further accelerated by exploiting redundancies. For fixed α and β as well as location-zero scale-one z, observe that several parts of (8) remain constant and only need to be evaluated once for all z. By removing these redundancies, we find that computation times decrease significantly, in particular those of the integrand, since the recursive Simpson quadrature requires a large number of function evaluations (a fact that is shared by most numerical integration methods). The resulting routine for the stable Paretian density is about 40 times faster than the naive implementation but equally robust and accurate. Moreover, the routine is also 8 times faster than the direct evaluation of the stable density found in John Nolan's STABLE 4.0 toolbox. In addition, that routine has an error, see Figure 1.

2.3.2 Further Methods of Density Approximation

As mentioned above, several ways have been proposed for computing the density of a stable Paretian variate. In addition to the direct approach in the previous section based on the real integral representation, we consider two further methods: First, the fast Fourier transform (FFT) based inversion of the characteristic function; and second, the spline-based, very fast approximation developed for the STABLE toolbox of John Nolan.



$x = 2$, $\beta = 0.2$, $\mu = 0$, $\sigma = 1$ $\qquad$ $x = 0$, $\beta = 0$, $\mu = 1$, $\sigma = 1$

Figure 1: Plots of the stable Paretian density as a function of the α-stable parameter in the neighborhood of $\alpha = 1.99$ for both the spline approximation, `stableqkpdf`, and the direct evaluation, `stablepdf`, as found in Nolan's STABLE 4.0 toolbox. Both panels illustrate a numerical discontinuity. This discontinuity indeed causes problems during optimization. For illustration, the remaining parameters are fixed to the values shown below each panel, but the problem arises for different values of x, β, μ and σ as well. Moreover, the problem is also found in the corresponding cdf routines.

The FFT approach first inverts the characteristic function based on a finite grid and then evaluates the pdf at the desired points by a subsequent interpolation. This approach is generally applicable and quite effective whenever the characteristic function is easily evaluated; see Mittnik et al. (1999); and Paolella (2007, Chapters 1 and 8). The accuracy of the approach (controlled by the grid size), however, is limited as the maximal grid size is restricted by memory constraints. For extreme events in the outer tail area in particular, the achievable accuracy in practice is not sufficient for likelihood optimizations as the routine may return the same likelihood at different evaluation points. Besides, there are estimation issues with dynamic grid sizes, which are often used for speed improvements. What turns out to be a convenient mechanism for reducing computation times becomes problematic in likelihood optimization as likelihood values become a function of the (varying) grid size. As a consequence, the estimation may not converge or, even worse, the estimation problem might not be identified (numerically) anymore.

In contrast, turning to Nolan's STABLE 4.0 toolbox (being closed source unfortunately), the spline approximation therein is more robust and, as to be expected for approximations, much faster. The increase in execution time is roughly a factor of 500 compared to the vectorized version of (8), though when computing the likelihood of the MixStable-GARCH model, this factor is only about 3 to 5, because of the time required for the matrix-based mixture GARCH filter. However, by construction, spline approximations will be correct to only a few significant digits, and can sometimes be problematic. Unfortunately, we also find numerical inconsistencies for $\alpha > 1.99$ as an error source in our estimations, see Figure 1. To circumvent these problems we rely on alternative methods (e.g., interpolation) in this region.

Our tests, based on real and simulated data, using (9) and Nolan's spline approximation suggest that the best compromise between speed and accuracy is a combination of both. As the fastest method, the spline approximation (with correction for the discontinuity mentioned above) lends itself well for an initial exploration of the search space, while the direct evaluation of (8) yields more reliable and accurate results. As such, we use Nolan's spline approximation to obtain initial estimates for the MixStable-GARCH model, which are then used as starting values in the subsequent re-estimation of the model based on the vectorized routine for (8). Results only based on the spline approximation were clearly inferior to those which are "polished" with a more accurate evaluation of the density.

2.3.3 Quantiles

In the empirical analysis below, we will additionally need the quantiles of the distribution from (3) in order to compute the VaR. If X follows a stable mixture random variable with pdf as in (3), then its cdf is just $\Pr(X \leq x) = \sum_{i=1}^{k} \omega_i F_S(x; \alpha_i, \beta_i, \mu_i, \sigma_i)$, where $F_S(x; \alpha, \beta, \mu, \sigma)$ is the cdf of the stable. The latter can be evaluated using the Gil-Pelaez (1951) inversion formula or the cdf integral expression of Zolotarev (1986) which avoids the need for complex numbers. For the case with $\alpha > 1$, Zolotarev's cdf expression is given by

$$F_S(z, \alpha, \beta) = \begin{cases} 1 - \frac{1}{\pi} \int_{-\tau}^{\pi/2} \exp\{-V(y; \alpha, \beta, z)\}\, \mathrm{d}y, & \text{if } z > \zeta, \\ \left(\frac{\pi}{2} - \tau\right)/\pi, & \text{if } z = \zeta, \\ 1 - F_S(-z, \alpha, -\beta), & \text{if } z < \zeta. \end{cases}$$

2.3.4 Optimization and Starting Values

Regarding the maximization of mixture likelihood functions, it appears beneficial to diversify among existing optimization methods for improving the chances of finding the global maximum given the inherently bumpy surface of such functions. We consider a combined optimization approach based on different search strategies from the rich set of unconstrained optimization techniques, including quasi-Newton and simplex methods. Using unconstrained optimization techniques, all constraints must be satisfied manually. We compute the mixture weights and the location parameters, respectively, from the $k-1$ estimated coefficients, by

$$\omega_i = \begin{cases} \hat{\omega}_i \left(1 - \sum_{j=1}^{i-1} \omega_j\right), & \text{if } i < k, \\ 1 - \sum_{j=1}^{k-1} \omega_j, & \text{if } i = k, \end{cases} \quad \text{and} \quad \mu_i = \begin{cases} \hat{\mu}_i, & \text{if } i < k, \\ -\left(\sum_{j=1}^{k-1} \omega_j \mu_j\right)/\omega_k, & \text{if } i = k, \end{cases}$$

where $\hat{\mu}_i$ and $\hat{\omega}_i$ are supplied by the optimization. Alternatively, $\boldsymbol{\omega}$ can be computed from the k coefficients by $\omega_i = \hat{\omega}_i / \sum_{j=1}^{k} \hat{\omega}_j$.[4] Box-constraints on single parameters, e.g., $1 < \alpha_i \leq 2$ and

[4] This, however, is numerically less efficient as it unnecessarily increases the dimension of the problem.

$0 \leq \omega_i \leq 1$, are satisfied by simple interval transformations. All estimations are terminated at a maximum of $50,000$ function evaluations, or, whenever the maximal change in the function value or the parameter vector is smaller than 10^{-4}. Numerical problems due to $\sigma_{i,t} \to 0$ or $\sigma_{i,t} \to \infty$ are addressed by enforcing $\underline{x} \leq \sigma_{i,t} \leq \overline{x}$, where $\underline{x}$ and $\overline{x}$, respectively, denote the smallest and largest (finite) machine number.

In addition to the above optimization technique and the joint estimation procedure in Section 2.3.2, we further improve upon the estimation quality by repeating the procedure with different starting values, and picking the best outcome in terms of the (augmented) likelihood value. Along with use of starting values based on previously estimated financial returns series, we also use values which are uniformly drawn from the allowed parameter space. For each estimation, at least three random starting vectors are considered. In consecutive estimations (such as used in our out-of-sample forecast exercises) we also include the previous best estimate. In all situations, the proposed estimation procedure converged irrespective of the starting value. In general, this would be remarkable, given the problematic nature of mixture likelihood functions, but is due to the use of the ALE estimation method discussed below in Section 3. Moreover, often (but not always), the various starting values led to the same optimum, further increasing our confidence that it is the global maximum.

3 Estimation of Mixture GARCH Models

It is well-known that the likelihood function of a mixture is potholed with singularities (infinite likelihood values), where single mixture components approach Dirac's delta distribution (see Kiefer and Wolfowitz, 1956; and Day, 1969). In the estimation of MixNormal-GARCH models, such *degenerated* states can be avoided by using Bayesian estimation procedures as devised in Ausín and Galeano (2007) and Bauwens et al. (2007). A potential drawback of these approaches, however, is their computational complexity and relatively high estimation time, along with the added complexity of augmenting those procedures from the normal to the stable (or other) distribution. To address these shortcomings, we devise an augmented likelihood function which can still be maximized with conventional optimization techniques but, unlike the usual likelihood, completely and elegantly avoids degenerated mixture estimates. As the problem of avoiding degenerated mixture estimates is a general problem known by many names, we will refer to it as the *mixture degeneracy problem*.

3.1 Augmented Likelihood Estimation

Augmented likelihood estimation (ALE) is based on the observation that degenerated components yield zero likelihood values for all observations which differ from their location parameter. As such, singularities can be prevented by avoiding zero likelihood values in all mixture components. This idea is implemented in the augmented likelihood framework by adding the geometric average likelihood value of each mixture component to the overall likelihood function. The additional k likelihood terms yield an infinite penalty whenever a mixture component degenerates. Thus, the ALE solves the mixture degeneracy problem without parameter constraints or penalty terms, but rather by shrinking the overall mixture likelihood function towards the component-wise likelihood functions.

Consider first estimation of an unconditional (no GARCH) mixture distribution. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)'$ denote the vector of model parameters, where $\boldsymbol{\omega}$ is the vector of mixture weights, and $\boldsymbol{\theta}_i$ represents the parameter vector of the density function, f_i, of the ith component. Then, the ALE takes the form of $\hat{\boldsymbol{\theta}}_{\mathrm{ALE}} = \arg\max_{\boldsymbol{\theta}} \tilde{\ell}\,(\boldsymbol{\theta}; \boldsymbol{\varepsilon})$, where the augmented log-likelihood function $\tilde{\ell}$ consists of the usual sum of log-likelihood values $\ell^{\star}$, as well as the logarithmic geometric means of the component-wise likelihood series $\bar{\ell}_i$, i.e.,

$$\begin{aligned}\tilde{\ell}\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right) &= \ell^{\star}\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right)+\sum_{i=1}^{k}\bar{\ell}_i\left(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}\right)\\ &= \sum_{t=1}^{T}\log\sum_{i=1}^{k}\omega_i f_i\left(\varepsilon_t;\boldsymbol{\theta}_i\right)+\sum_{i=1}^{k}\frac{1}{T}\sum_{t=1}^{T}\log f_i\left(\varepsilon_t;\boldsymbol{\theta}_i\right). \end{aligned} \tag{10}$$

Assuming (as we subsequently do) that the true mixture process is free of degenerated components and its standard likelihood function has a consistent root (compare Kiefer, 1978), the ALE is consistent, because, by dividing by T, the contribution to $\tilde{\ell}$ from the additional k terms becomes negligible in the limit as $T\to\infty$ for all mixtures without degenerated components.[5]

3.2 Augmented Mixture GARCH Estimation

The augmented likelihood methodology above is tailored to unconditional mixtures and, thus, cannot prevent singularities in (time-) conditional mixture models. In k-component mixture GARCH models, for instance, singularities occur at all time points t, $t=1,\dots,T$, and all components i, $i=1,\dots,k$, where $\sigma_{i,t}=0$ and $\mu_i=\varepsilon_t$. We refer to this problem as the *local degeneracy problem*. Assuming that the true mixture GARCH process does not possess singular or near-singular components at any point in time, we devise two solutions for the local degeneracy problem based on the augmented likelihood methodology.

The first solution is built upon the ALE by removing singularities from the (mixture) likelihood function using a lower bound on the GARCH constant(s),

$$\gamma_{i,0}>c>0, \quad \text{for all } \ i=1,\dots,k. \tag{11}$$

This parallels the idea in Hathaway (1985) in the simple (unconditional) normal mixtures, and is extended in Tanaka (2009). Without local degeneracy, the ALE remains feasible and the mixture GARCH degeneracy problem can be solved as in the unconditional case. The *restricted* ALE (RALE) for the MixStable-GARCH model is given by $\hat{\boldsymbol{\theta}}_{\mathrm{RALE}}=\arg\max_{\boldsymbol{\theta}}\tilde{\ell}\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right)$ where (11) is satisfied and (10) becomes

$$\tilde{\ell}_S\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right)=\sum_{t=1}^{T}\log\sum_{i=1}^{k}\omega_i f_S\left(\varepsilon_t;\alpha_i,\beta_i,\mu_i,\sigma_{i,t}\right)+\sum_{i=1}^{k}\frac{1}{T}\sum_{t=1}^{T}\log f_S\left(\varepsilon_t;\alpha_i,\beta_i,\mu_i,\sigma_{i,t}\right).$$

A shortcoming of this approach is the necessity of having to choose the tuning parameter c, and doing so in such a way that, as a function of sample size T, the sequence c_T yields a consistent estimator. Extensive studies with the financial data used in this report, and with simulated data mimicking such finance data, yield that $c=0.01$ works well in the sense that, while (11) is occasionally binding, the parameter estimates barely differ from those obtained by the EALE method discussed next.

Our second solution to the local degeneracy problem is more general and can be applied to any conditional mixture model. Let $\boldsymbol{\ell}\in\mathbb{R}^T$ denote a vector of (finite) log-likelihood values. Based on the observation that the sample geometric mean of the likelihood values

$$\widehat{\mathbb{E}}_{\mathrm{geo}}(\boldsymbol{\ell})=\exp\left\{\frac{1}{T}\sum_{t=1}^{T}\ell_t\right\}=\exp\left\{\bar{\ell}\right\}$$

grows at a slower rate than the sample variance-like quantity

$$\widehat{\mathbb{V}}_{\mathrm{geo}}(\boldsymbol{\ell})=\frac{1}{T}\sum_{t=1}^{T}\left(\exp\left\{\ell_t\right\}-\exp\left\{\bar{\ell}\right\}\right)^2$$

[5] An example of a non-consistent ALE (or MLE) can easily be constructed by imposing inappropriate parameter constraints. Indeed, this is the reason why the RALE (below) is not always a consistent estimator.

if some likelihood values move towards infinity, i.e.,

$$\frac{\widehat{\mathbb{E}}_{\text{geo}}(\boldsymbol{\ell})}{\widehat{\mathbb{V}}_{\text{geo}}(\boldsymbol{\ell})} \to 0 \ , \quad \text{if } \widehat{\mathbb{E}}_{\text{geo}}(\boldsymbol{\ell}) \to \infty,$$

we propose the component-wise penalty term, $\check{\ell}_i$,

$$\check{\ell}_i\left(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}\right) = \log\left(1 + \frac{1}{T}\sum_{t=1}^{T}\left(\exp\left\{\ell_{i,t}\right\} - \exp\left\{\bar{\ell}_i\right\}\right)^2\right), \tag{12}$$

where $\ell_{i,t} = \log f_i\left(\varepsilon_t;\boldsymbol{\theta}_i\right)$. The component-wise incorporation of (12) into (10) consequently removes all local singularities. Moreover, as (12) is purely likelihood-based, the resulting estimator is also free of direct parameter constraints. The *extended* ALE (EALE) is given by $\hat{\boldsymbol{\theta}}_{\text{EALE}} = \arg\max_{\boldsymbol{\theta}} \tilde{\tilde{\ell}}\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right)$ where

$$\tilde{\tilde{\ell}}\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right) \;=\; \ell^{\star}\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right) \;+\; \sum_{i=1}^{k}\bar{\ell}_i\left(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}\right) \;-\; \sum_{i=1}^{k}\check{\ell}\left(\boldsymbol{\theta}_i;\boldsymbol{\varepsilon}\right).$$

Similar to the ALE, the additional terms vanish as $T \to \infty$, so that the EALE is consistent. The price to pay is that the penalty in (12) introduces a source of bias as the density function is penalized for (too) large likelihood values, thus implicitly transferring mass to the tail area. For the case of the MixStable-GARCH model, the expression specializes to

$$\tilde{\tilde{\ell}}_S\left(\boldsymbol{\theta};\boldsymbol{\varepsilon}\right) = \sum_{t=1}^{T}\log\left(\sum_{i=1}^{k}\omega_i L_{i,t}^{S}\right) + \sum_{i=1}^{k}\left\{\log g_i - \log\left(1 + \frac{1}{T}\sum_{t=1}^{T}\left(L_{i,t}^{S} - g_i\right)^2\right)\right\},$$

where $L_{i,t}^{S} = f_S\left(\varepsilon_t;\alpha_i,\beta_i,\mu_i,\sigma_{i,t}\right)$ and $g_i = \left(\prod_{t=1}^{T} L_{i,t}^{S}\right)^{1/T}$.

Both RALE and EALE yielded essentially the same maximized likelihood values in all runs we conducted based on real and simulated data. (It is important to emphasize that the actual, and not the augmented, likelihood is being referred to here.) While there was a slight preference for the EALE over the RALE in terms of actual likelihood value, both resulted in essentially identical out-of-sample forecasts. The local degeneracy problem for MLE, RALE and EALE is further illustrated in Figure 2. For all data generating processes we have studied, the results look qualitatively the same: while RALE and EALE never yield degenerated estimates (all likelihood values are finite), the standard MLE frequently results in degenerated estimates (infinite likelihood values). It is due to the imperfection of global optimization methods that, first, the MLE does not always degenerate, and second, that degenerated estimates occur less frequently for larger sample sizes, as can be seen by comparing the left and right panels in Figure 2.

3.3 Small Sample Properties

Simulation studies were conducted in order to assess the distributional properties of the estimators, partly to compensate for the fact that the asymptotic properties are elusive at present. The parameters of the data generating processes we consider are calibrated to actual finance data. The first analysis is based on the estimation error $\epsilon_{i,j}^{P} = \hat{\theta}_{i,j} - \theta_j$, $i = 1,\ldots,N$, $j = 1,\ldots,M$, where $\boldsymbol{\theta}$ is the parameter vector of the true process, M is the total number of parameters, and $N = 1000$ the number of simulations. To address the label switching problem, parameter vectors are sorted by mixture weights. Table 1 shows the results of two simulation studies for the MixStable(k, g) model.[6] Two measures are reported. The first is the usual root mean squared

[6] More detailed results and information on computational aspects are available from the authors upon request.

error of the parameters (P-RMSE), $\sqrt{\sum_i (\epsilon^P_{i,j})^2/N}$.[7] The second measure is the inter-quantile range (IQR).[8] As expected, all error measures decrease as the sample size increases.

We now turn to the second analysis. The above parameter error diagnostic needs to be augmented as a basis for the analysis for two reasons: First, mixture processes can often be mimicked quite accurately by processes with fairly different parameter vectors. Second, the sorting of mixture components may lead to false results, e.g., if some of the true mixture weights are close. Hence, we consider two proxy measures that reflect the main characteristics of mixture GARCH processes. The first is based on the quantiles of the unconditional distribution of the process and the second uses the autocorrelation function (acf) of the absolute returns. (In order to compare the sample values to their theoretical counterparts, the latter are obtained via simulation of the true process using one million observations.) The corresponding RMSE measures (Q- and A-RMSE) are given by the square root of the mean squared

$$\epsilon^Q_{i,j} = \text{quantiles}(\hat{\theta}, \boldsymbol{\zeta})_{i,j} - \text{quantiles}(\theta, \boldsymbol{\zeta})_{0,j} \text{ and } \epsilon^A_{i,j} = \text{acf}(\hat{\theta}, \rho)_{i,j} - \text{acf}(\theta, \rho)_{0,j},$$

respectively, where we evaluate the sample quantile function at the probability levels $\boldsymbol{\zeta} = (0.01, 0.1, 0.2, \ldots, 0.9, 0.99)$ and compute the sample acf with lag order $\rho = 100$.

For comparison we also report the performance of the standard MLE (although it is ill-defined for mixtures). As expected, the MLE often results in estimates for which at least one component explains only a few observations. Such mixture components are typically either degenerated or correspond to local optima where fewer mixture components explain the data (in the sense of almost zero-valued mixture weights) than actually being estimated. Clearly, RALE and EALE outperform the MLE here, but also suffer from (non-singular) local optima, though substantially less. Similarly, it is evident that the EALE results in non-stationary estimates more often than the RALE. The reason is that in order to match a given unconditional volatility, an increase in the GARCH constant must be offset by a decrease in the remaining GARCH parameters, thus moving them away from the non-stationarity border. This shortcoming of the EALE, however, vanishes as the sample size increases. In contrast, by looking at the IQR values based on all estimates, Table 1 indicates that RALE and EALE perform very similarly; in the two-component case the RALE slightly outperforms the EALE, while it is the other way around in the three-component case. Although the RALE tends to be slightly faster, we opt for the EALE in the following, as it has the appealing advantage of avoiding direct parameter constraints and is therefore an unrestricted solution of the mixture degeneracy problem.[9]

4 Univariate Empirical Results

Our empirical analysis covers the major international equity indices DAX 30, S&P 500, DJIA 30, NIKKEI 225 and NASDAQ COMPOSITE (20 years, dating back from July 7th, 2009; resulting in a sample size of 2609) as well as the exchange rates JPY/EUR and USD/EUR (10 years, dating back from July 7th, 2009; resulting in a sample size of 1304). All results are based on percentage log returns, $\varepsilon_t = 100\,(\log p_t - \log p_{t-1})$, where p_t is the daily closing index price at time t. We study MixNormal(k, g), MixGED(k, g) and MixStable(k, g) models, where

[7] We exclude those runs (estimated parameter vectors) such that one or more components had an excessively small weight (in particular, when $\omega_i T < 10$, e.g., the ith component explains less than 0.1% of the observations when using $T = 1000$). We do this to account for the RMSE sensitivity to numerical outliers in the parameter estimates of such components.

[8] Unlike the P-RMSE, the IQR is based on all runs, as it is not affected by outliers.

[9] Regarding computation time: On an Intel i7-2600K quad-core processor at 4.2Ghz a single estimation (EALE) of the MixStable$(4, 4)$ model based on 1000 data points (running in Matlab R2010a on a single core of the cpu) takes up to four minutes using Nolan's fast spline approximation and about four times the amount of time using the vectorized version of Zolotarev's integral expression.

$(k, g) \in \{(2,2),(3,2),(3,3),(4,4)\}$.[10] To prevent overfitting, shape parameters are restricted to be identical.[11]

Cutting to the chase, the best performing models (in terms of both in-sample fit and, more importantly, out-of-sample performance) are all in the MixStable class. In-sample, the A^2MixStable$(2,2)$ model is favored by the BIC, while out-of-sample, the A^2MixStable$(2,2)$, as well as the A^2MixStable$(3,2)$ and A^2MixStable$(3,3)$, perform well. It is noteworthy that the A^1MixStable$(4,4)$ model, despite its high parametrization, performs overall best in terms of the uniformity of the predictive cdf values. Nevertheless, the A^2MixStable$(3,2)$ model might be preferred because of its relatively parsimonious parametrization (which also implies faster estimation). It also performed slightly better than the A^1MixStable$(4,4)$ for the VaR comparisons at the lower probability levels.

4.1 Choice of δ

We investigate the influence of the GARCH power parameter δ, for which $\delta < \alpha$ must be satisfied in order to ensure the existence of the MixStable-GARCH process. In particular, given the flexibility and richness of the model, we have confirmed for numerous data sets and sample sizes that, relative to the other parameters, the likelihood is relatively flat in δ. This can lead to exacerbated estimation problems (in addition to requiring dynamically imposed constraints, as opposed to simple box constraints, during estimation), and implies that just setting δ to a compromise value will not lead to appreciably poorer (and could possibly lead to slightly better) forecasts. We find that the choice of $\delta = 1$ is not only adequate, but also conveniently satisfies the $\delta < \alpha$ constraint.

Consider the profile log-likelihood function, $\ell_{\mathrm{p}}(\delta; \boldsymbol{\varepsilon}) = \max_{\boldsymbol{\eta}} \ell^{\star}(\boldsymbol{\eta}; \boldsymbol{\varepsilon}, \delta)$, where $\ell^{\star}$ refers to the standard likelihood function, $\boldsymbol{\varepsilon}$ is the vector of asset returns, and $\boldsymbol{\eta}$ denotes the entire parameter vector except for δ. Table 2 shows the (in-sample) profile likelihood estimates using the EALE. (We show only the 4-component models under study as the results for different numbers of mixture components are qualitatively the same.) The optimal choice of δ is between one and two, with a tendency towards one. In addition, Table 3 shows that $\delta = 1$ (compared to $\delta = 2$) also improves the out-of-sample density forecasts across models for the majority of data sets in this paper. It also yields lower risk prediction errors in terms of the cumulated root mean squared error over the important VaR level up to 5%.

Our result is also in line with similar studies for simpler GARCH models. In particular, when fitting an asymmetric power GARCH model (APARCH) process with Student's t innovations to 10 national stock indices and the MSCI world index, Brooks et al. (2000) find that the power parameter is significantly different from unity for only one series, whereas it is significantly different from two in nine cases. Similarly, Giot and Laurent (2003) apply a skewed Student's t APARCH model to three national stock markets and three individual stocks. For all but one of the stocks, the estimated δ is rather close to (and statistically indistinguishable from) unity, leading the authors to conclude that "*instead of modeling the conditional variance (GARCH), it is more relevant to model the conditional standard deviation*". More recently, Lejeune (2009) obtains similar results for the S&P 500 and NASDAQ data. Clearly, when working with stable

[10] Several models with $k > 4$ were also considered; information criteria never favored them, out-of-sample forecasts were roughly comparable to the $k = 4$ case but increasingly mixed for higher k, and estimations resulted more frequently in estimates with mixture weights close to zero for the additional components. We do not report the results.

[11] By relaxing the equality of the shape parameters constraint, the resulting estimates show an undesired property: While very few mixture components explain the majority of the data with reasonable parameter estimates, the majority of the mixture components maximizes the likelihood based on corner solutions for their shape parameters, which is a classic indicator of overfitting. Using 1000 observations, the MixGED(k, g) and MixStable(k, g) are frequently overfitted, while for larger data sets, the overfitting vanishes, as would be expected, and for the full DJIA sample, MixGED(k, g) estimates with free shape parameters are found to be in line with those reported in Rombouts and Bouaddi (2009).

distributions, it no longer makes sense to speak of standard deviations, but rather the scale terms in each of the k components.

Observe that setting $\delta = 1$ implies that the MixStable-GARCH model no longer nests the MixNormal(k, g) model in (1) and (2), but rather a variant of it, in which the exponent of 2 is replaced by 1.

4.2 In-Sample Fit

For assessing in-sample properties, we fit the MixNormal-, MixGED- and MixStable-GARCH models to the seven financial return series under study using the extended likelihood estimator (EALE) from Section 3.2. Exemplarily, Table 4 shows the parameter estimates of the two MixStable-GARCH models for the DJIA return data.

Table 5 shows various in-sample statistics of all models and data sets under study. Out of necessity, the most general model, A^1MixStable$(4, 4)$, yields the highest likelihood value for all seven data sets. (This otherwise clear result unfortunately does not carry over to the investigation of out-of-sample forecasts.) As expected, the BIC measure favors less densely parametrized models, with an overall tendency towards the A^2MixStable$(2, 2)$ model. We focus on the BIC results as the literature on mixture models provides some theoretical and empirical support for its appropriateness and good performance, in particular for selecting the number of mixture components (see, e.g., Keribin, 2000; Francq et al., 2001; and Frühwirth-Schnatter, 2006, Ch. 4). On the contrary, AIC results are mixed with no clear pattern.

4.3 Comparison of Forecasting Performance

While a model as flexible as the MixStable(k, g) should be expected to provide an excellent in-sample fit to virtually any return series compared with more traditional GARCH-type models, the concern remains as to whether the relatively large parametrization, the nontrivial computational aspects of the stable density, and the degeneracy issue associated with mixtures warrant its use. To judge this, we compare the empirical performance of the one-step-ahead predictive cdfs across models using tests for uniformity (see below) as well as concentrating on the left tail, using probability values typical for VaR calculations.

In particular, for a given target probability, λ, typically chosen between 1% and 10%, the VaR delivers an upper bound on losses such that it will be exceeded with probability λ. Conditional on the information given up to time $t-1$, the VaR for period t of one unit of investment is the negative λ-quantile of the conditional return distribution, i.e., $\mathrm{VaR}_{t|\mathcal{F}_{t-1}}(\lambda) = -\inf_x\{x \in \mathbb{R} : \Pr(\varepsilon_t \leq x \mid \mathcal{F}_{t-1}) \geq \lambda\}$ for $0 < \lambda < 1$, where ε_t is the return on an asset or portfolio in period t. In our case, with a continuous, strictly monotone increasing predictive cdf $F_{t|\mathcal{F}_{t-1}}$, we have $\mathrm{VaR}_{t|\mathcal{F}_{t-1}}(\lambda) = -F^{-1}_{t|\mathcal{F}_{t-1}}(\lambda)$. Observe that, while the VaR is considered an inferior risk measure compared to expected shortfall (see, e.g., Dowd, 2005 and the references therein), it is still immensely popular, and the computation of the more sophisticated expected shortfall requires accurate calculation of the VaR.

The choice of the window size is a tuning parameter chosen to maximize the quality of future risk or density forecasts of the particular set of assets under study. It is not necessarily the case that more observations are better, because, with certainty, the proposed model differs in some way from the true data generating process, which itself is likely to not be strictly stationary over long periods of time. A model is possibly, however, a good approximation to reality for short periods of time, though using too short a window results in high variance of the parameters and thus inferior forecasts. To negotiate this bias/variance tradeoff, some experiments with the MixStable-GARCH model and two of the data sets under study indicate that use of a rolling window of sample size 1000 is superior to use of either 500 or 2000. While this value could be optimized further by conducting dedicated experiments per dataset, we use 1000 in all of the following.

For all models considered, we re-estimate the model parameters every 20 trading days (about once a month), so that each estimation contains 2% of new data. Our analysis is based on the realized predictive cdf values obtained from evaluating the one-step-ahead cdf forecasts at the realized returns. If the model is correct, it is well-known that these are uniformly distributed.

Let $\hat{p}_t = \hat{F}_{t|\mathcal{F}_{t-1}}(\varepsilon_t; \hat{\boldsymbol{\theta}}_{t-h})$, $t = 1, \ldots, N$, be the sequence of realized predictive cdf values, noting that, for each t, the parameter vector is estimated using information (in this case, just the past returns) up to and including time $t - h$, where h is a value in $\{1, 2, \ldots, 20\}$, but the entire return series up to time $t - 1$ is used in the model filter. Finally, this predictive cdf is evaluated at the actual return at time t. Denote the collection of these N values as vector $\hat{\boldsymbol{p}}$. Further let $\hat{\boldsymbol{p}}^{[s]}$ denote the sorted vector, $\hat{p}_1^{[s]} \leq \hat{p}_2^{[s]} \leq \cdots \leq \hat{p}_N^{[s]}$. The Anderson-Darling (AD) and Cramér-von Mises (CM) test statistics are given respectively by

$$\mathrm{AD} = -N - \sum_{i=1}^{N} \frac{2i-1}{N} \left(\log(\hat{p}_i^{[s]}) + \log(1 - \hat{p}_{N-i+1}^{[s]}) \right)$$

and

$$\mathrm{CM} = \frac{1}{12N} + \sum_{i=1}^{N} \left(\frac{2i-1}{2N} - \hat{p}_i^{[s]} \right)^2 .$$

In addition, we provide test statistics for the Kolmogorov-Smirnov (KS) test for uniformity, as well as the Jarque-Bera (JB) and Shapiro-Wilk (SW) tests for normality after applying the inverse normal cdf transform. We test for serial correlation in $\hat{\boldsymbol{p}}$ (as a proxy for the iid property) and report Ljung-Box (LB) test statistics,

$$\mathrm{LB} = N(N+2) \sum_{i=1}^{m} \frac{\hat{\rho}_i^2}{N-i},$$

where $\hat{\rho}_i$ is the ith autocorrelation from the ith sample autocorrelation function.

Tables 6 and 7 show the results. It is important to note that here, we are testing the prediction quality over the whole support of the distribution, and not just the left tail (as we do below, for directly testing the quality of value at risk predictions). Except for DAX and NASDAQ, AD and CM are clearly in favor of the stable models with a strong preference for the A^1MixStable-GARCH model. Results for KS and LB are less clear but also pro stable in four out of seven cases. Similar results are obtained for both normality tests, though the JB test appears to favor the A^2MixStable-GARCH model.

We also consider VaR measures dedicated to the left tail, as these are of possibly even greater interest from a risk management perspective. Table 8 shows the empirical coverage probabilities (as percentages) for the 1%, 5% and 10% VaR levels along with p-values indicating the severeness of potential risk underestimation. The results for 5% and 10% are mixed, though at the 1% VaR level (arguably the most important VaR level in risk management applications), the A^2MixStable(k, g) model clearly outperforms the other models for most data sets under study.

For further investigations of the VaR prediction quality, we adopt a simple quality measure based on the coverage error over the VaR levels up to $100\lambda\%$, see Kuester et al. (2006). The measure calculates the deviation between predictive cdf and uniform cdf and, thus, captures the excess of percentage violations over the VaR levels, where the deviation is defined as $100(F_U - \hat{F}_e)$ with F_U being the cdf of the standard uniform random variable and $\hat{F}_e$ referring to the empirical cdf formed from $\hat{\boldsymbol{p}}$. Building upon this metric we report the integrated root mean squared error (IRMSE) over the left tail up to the maximal VaR level of interest. The IRMSE employed herein is closely related to the CM statistic but with the sum truncated at $h = \lceil \lambda N \rceil$, i.e.,

$$\mathrm{IRMSE} = \sqrt{\frac{1}{h} \sum_{i=1}^{h} \left(100 \frac{2i-1}{2N} - 100 \hat{p}_i^{[s]} \right)^2 } .$$

The results in Table 9 confirm the superiority of the stable models in five out of seven cases at the 1% level, and also at the 10% VaR level.

Finally, we investigate the hit sequence of realized predictive VaR violations,

$$v_t = \mathbf{1}_{\varepsilon_t \leq \hat{q}_t}, \quad \hat{q}_t = \widehat{\mathrm{VaR}}_{t|\mathcal{F}_{t-1}}(\lambda), \tag{13}$$

where $\mathbf{1}$ is the indicator function. Under the null of correct conditional coverage, the v_i are iid Bernoulli(λ). From this sequence, the test statistic $\mathrm{LR}_{\mathrm{CC}} = \mathrm{LR}_{\mathrm{UC}} + \mathrm{LR}_{\mathrm{IND}}$ is computed, as proposed in Christoffersen (1998), where $\mathrm{LR}_{\mathrm{UC}}$ and $\mathrm{LR}_{\mathrm{IND}}$ test for unconditional coverage and independence, respectively. As can be seen from Table 10 for the 1% VaR level, the tendency of MixStable(k, g) (in particular of A^2MixStable-GARCH) to outperform MixNormal(k, g) and MixGED(k, g) is also corroborated by the Christoffersen test.

5 ICA-MixStable-GARCH

A direct generalization of the MixNormal-GARCH model to the multivariate setting with D assets has been investigated by Bauwens et al. (2007) and Haas et al. (2009), the latter model allowing for asymmetries. While of value for a small number of assets, those models will not be practical for even modest portfolios, let alone large ones. In addition, attempting to extend that model to support the multivariate stable distribution is not trivial (but see Lombardi and Veredas, 2009; Bonato, 2011; and the references therein).

A multivariate distribution with mix-Stable GARCH marginals can be constructed via use of independent components analysis (ICA); see, e.g., Hyvärinen et al. (2001). Crucially, the resulting multivariate distribution is such that the distribution of a linear combination (as needed to conduct portfolio optimization) is tractable. The method assumes a set of non-Gaussian distributed independent random variables of which linear combinations in the form of time series have been observed. The goal is to recover the original independent random vectors of time series, called the independent components. Once the mixing matrix is known (estimated), the independent components can be modeled and forecasted independently (by, for example, a GARCH-type model). The related concept of conditionally uncorrelated components is discussed in Fan et al. (2008).

An application of the popular iterative FastICA algorithm can be found in Broda and Paolella (2009), where the method is used in a portfolio allocation exercise to estimate the independent components (driven by generalized hyperbolic innovations) of the 30 constituents of the Dow Jones Industrial Average index. To be precise, the ICA variant employed therein maximizes the conditional heteroskedasticity of the independent components. We refer to this specific version as CHICA (Conditionally Heteroskedastic ICA). Unfortunately, the CHICA method has the drawback of requiring finite fourth moments, which is not fulfilled in our setting. However, to the best of our knowledge, the only ICA method dedicated to (time-correlated) stable driven data is found in Fabricius et al. (2001). It is all the more remarkable that in our extensive simulation studies, the method is outperformed by the CHICA method.

Briefly, the CHICA method is a two-step procedure, separating the estimation of the correlation structure from that of the univariate dynamics; details are given in Chen et al. (2006) and Broda and Paolella (2009). A first step is to estimate the expectation of the D assets and obtain the matrix of de-meaned returns $\boldsymbol{Y} \in \mathbb{R}^{D \times T}$. There are several ways of doing this. One could just use the sample mean for each series, though given the fat-tailed nature of the data, this is not efficient. A trimmed mean, or even the median, might be superior, or perhaps best, use of the location term jointly estimated with, say, a MixStable-GARCH model. Whichever is used, further improvement might be realized by (i) estimating it using weighted likelihood, with relatively more weight on more recent observations (see Paolella and Steude, 2008; Broda and Paolella, 2011; and the references therein) and/or (ii) shrinking the D values towards, say, zero.

In the second step, the CHICA method is applied to the residual series $\boldsymbol{Y}$ estimating the mixing matrix $\boldsymbol{A} \in \mathbb{R}^{D \times D}$ that best separates the signals maximizing their GARCH effects.

Given the mixing matrix $\boldsymbol{A}$, the D independent components, $\boldsymbol{X} \in \mathbb{R}^{D\times T}$, are obtained by $\boldsymbol{X} = \boldsymbol{A}^{-1}\boldsymbol{Y}$. The third (and only time-consuming) step consists in the estimation and prediction of the D (univariate) independent components based on the MixStable-GARCH model. Observe that the three above steps (in particular, the last one) do not need to be repeated for different portfolio weights. This is crucial if such weights are to be optimized. Note also that each of the D components is endowed with its own, estimated, MixStable-GARCH model, and so each has its own tail index α. This can be contrasted with attempts using the multivariate stable distribution, which has only a single α for all assets.

The distribution of the weighted sum can be computed via standard inversion methods applied to the characteristic function (cf) of the convolution. The stable mixture distribution has a simple expression for its cf, whereas many of the *ad hoc* distributions used in this context, such as GED, do not have such forms, thus precluding their use. (As an aside perhaps worth mentioning, the Student's t distribution is applicable. Let $Z \sim t_\nu$ with zero location and unit scale, and $\nu \in \mathbb{R}_{>0}$. The characteristic function of Z is

$$\varphi_Z(t;\nu) = \frac{\nu^{\nu/4}|t|^{\nu/2}}{2^{\nu/2-1}\Gamma(\nu/2)} K_{\nu/2}\left(|t|\sqrt{\nu}\right), \tag{14}$$

where K_v is the modified Bessel function of the third kind with index v, $v > 0$. Result (14) is stated (without reference or derivation) in the reference work of Kotz and Nadarajah (2004, p. 40), while Platen and Heath (2006, p. 37) and Seneta (2004, p. 186) note that it is attributed to Simon R. Hurst, in an unpublished article in 1995 and his dissertation in 1997. Because of the Bessel function, its inversion will be far slower than the stable cf, but it indeed could be used in this context.)

Suppressing the time index for readability, the return R on a portfolio of D assets is a weighted sum of the individual asset returns, and consequently (via the matrix $\mathbf{A}$) also a weighted sum of the independent components X_j with associated weights b_j, say. The distribution of this weighted sum can be computed from its characteristic function (cf), which, from independence, factors into the product of the cfs of the individual components. In our context, each X_j is a k-component stable mixture, with pdf given by $f_{X_j}(x) = \sum_{i=1}^k \omega_{j,i} f_S(x;\alpha_{j,i},\beta_{j,i},\mu_{j,i},\sigma_{j,i})$. As the cf of a stable random variable with tail index α, asymmetry parameter β, location μ and scale σ is $\exp\{\imath\mu t - \sigma^\alpha |t|^\alpha (1 - \imath\beta \operatorname{sign}(t)\tan(\alpha\pi/2))\}$, $\imath$ denoting the imaginary unit, the cf of X_j is given by

$$\varphi_{X_j}(t) = \sum_{i=1}^k \omega_{j,i} \exp\left\{\imath\mu_{j,i}t - \sigma_{j,i}^{\alpha_{j,i}}|t|^{\alpha_{j,i}}\left(1 - \imath\beta_{j,i}\operatorname{sign}(t)\tan\left(\alpha_{j,i}\frac{\pi}{2}\right)\right)\right\},$$

from which a simple expression for $\varphi_R(t)$ follows. Standard inversion methods can then be applied to φ_R to calculate its pdf or cdf; see, e.g., Paolella (2007, Chapter 1) for details.

For MixStable(k,g), the model lends itself to the computation of portfolio risk measures such as expected shortfall (ES) or VaR due to its simple cf expression. (Interestingly, as shown in Sy (2006), the VaR is a coherent risk measure for portfolios of independent stable Paretian distributed assets, but only if the tail exponent α is the same for all assets.) For continuous return R, the (relative) expected shortfall at $100\lambda\%$ level, $0 < \lambda < 1$, is

$$\mathrm{ES}_R^\lambda = -\mathbb{E}[R|R<q] = -\frac{1}{\lambda}\mathbb{E}[R\mathbf{1}_{R<q}],$$

where $\mathbf{1}$ is the indicator function and $q = -\mathrm{VaR}_R(\lambda)$, the $100\lambda\%$ quantile.

If R were just a mixture of stable distributions, one could compute $\mathbb{E}[R\mathbf{1}_{R<q}]$ via the real integral expression for the stable ES in Stoyanov et al. (2006) combined with the general results for the ES of mixtures in Broda and Paolella (2011). However, as R is a convolution of mixtures of stables, this is not applicable. Neither is the general result of Kim et al. (2009a), which requires that the cf be finite in a strip containing the real axis, which is not the case for stable

distributions and their mixtures. To compute $\mathbb{E}[R\mathbf{1}_{R<q}]$, we use the result from Broda (2011) which applies more generally than that given in Kim et al. (2009a),

$$\mathbb{E}[R\mathbf{1}_{R<q}] = \frac{\varphi_R'(0)}{2\imath} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\mathrm{e}^{-\imath tq}\varphi_R'(t)\right]\frac{\mathrm{d}t}{t}, \tag{15}$$

where φ_R' is the first derivative of the characteristic function of R, as derived in Appendix A.

Taken together, the computational ingredients are now available to do portfolio optimization in return/ES space, using ICA with the independent components modeled as MixStable processes. The resulting portfolio will, by necessity, have Pareto-like tails, no matter how many assets are involved, unlike in Broda and Paolella (2009), for which, as it uses the normal inverse Gaussian distribution for the components, eventually as D increases, a central limit theorem effect will kick in, and the results will not differ from using a standard Markowitz approach. Given the number of tuning parameters and dependence on the choice of data, the assessment of the performance of the proposed method for portfolio allocation deserves a separate study, and is not pursued here.

6 Conclusions and Future Research

An open question in financial econometrics is the suitability of infinite-variance distributions for modeling the unconditional or conditional distribution of asset returns. As discussed at length in the introduction, there is an ongoing debate regarding the maximally existing moment of asset returns. In this paper, we take the stance, as argued by numerous researchers in the stable Paretian field, that the stable Paretian distribution has many positive characteristics for modeling asset returns that outweigh its potential shortcomings. Indeed, we find that a model incorporating mixtures, GARCH, and underlying stable distributions yields a complex but coherent and statistically well motivated model which, based on extensive empirical exercises, delivers both excellent in-sample fit and, most relevantly, admirable out-of-sample forecasting results. In particular, the A^2MixStable$(3,2)$, A^2MixStable$(3,3)$ and A^1MixStable$(4,4)$ models performed best out-of-sample.

Various model extensions suggest themselves for future consideration. There could be value in generalizing the law of motion for the scale terms along the lines of the asymmetric GARCH models in Ding et al. (1993) or Alexander and Lazar (2009). The incorporation of a Markov switching structure, as done in Haas et al. (2004a) and Bauwens et al. (2010), or extension to time-varying component weights, as recently investigated by Bauwens and Storti (2009), could also (and possibly in addition) be entertained.

Finally, as mentioned in Footnote 2, the tempered stable distribution could be considered in place of the stable Paretian, which would address the non-summability issue, and, given its existence of absolute moments, possibly be of value for option pricing; see Poirot and Tankov (2006), Mercuri (2008) and Kim et al. (2009b) for option pricing with a tempered stable distributional assumption, and Badescu et al. (2008) and Rombouts and Stentoft (2009) for option pricing with the mixed normal distributional assumption.

A First Derivative of the Characteristic Function of a Weighted Sum of Mixed Stable Random Variables

For evaluating the ES of a portfolio of mixed stable distributed asset returns based on (15), we derive the first derivative of the cf of sum of mixtures of stable Paretian random variables in the general case of asymmetric stable random variables and unrestricted portfolio weights. Similar to the portfolio of D assets in Section 5, with summation weights $\boldsymbol{b} \in \mathbb{R}^D$, the portfolio sum is given by $R = \sum_{j=1}^{D} b_j X_j$, where the X_j are independent k-component mixtures of stable random variables,

$$X_j \sim \text{MixStable}(\boldsymbol{\omega}_j, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j), \quad j = 1, \dots, D,$$

with mixture weights $\boldsymbol{\omega}_j \in (0,1)^k$, $\sum_i \omega_{j,i} = 1$, tail indices $\boldsymbol{\alpha}_j \in (0,2]^k$, skewness coefficients $\boldsymbol{\beta}_j \in [-1,1]^k$, location parameters $\boldsymbol{\mu}_j \in \mathbb{R}^k$ and scale terms $\boldsymbol{\sigma}_j \in \mathbb{R}^k_{>0}$. The cf of R is straightforward to calculate and takes the form of

$$\varphi_R\left(t; \boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\right) = \prod_{j=1}^{D} \sum_{i=1}^{k} \omega_{j,i} \varphi_P\left(t; b_j, \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i}\right), \tag{16}$$

where

$$\varphi_P\left(t; b, \alpha, \beta, \mu, \sigma\right) = \varphi_W\left(t; \alpha, \beta, b\mu, b\sigma\right) = \exp\left\{-\left|\sigma b t\right|^{\alpha}\left(1 - \imath\beta \operatorname{sgn}\left(bt\right) \tan\left(\alpha\frac{\pi}{2}\right)\right) + \imath\mu b t\right\},$$

with $W(\cdot, \alpha, \beta, \mu, \sigma)$ being a stable Paretian random variable with tail index α, asymmetry β, location μ, and scale σ. Straightforward but tedious algebra shows that the first derivative of (16) is

$$\varphi_R'\left(t; \boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\right) = \sum_{i=1}^{D}\left(s_i' + \sum_{j=1, j\neq i}^{D} s_j\right),$$

where

$$s_j = \sum_{i=1}^{k} \omega_{j,i} \varphi_P\left(t; b_j, \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i}\right), \quad j = 1, \dots, D,$$

$$s_j' = \sum_{i=1}^{k} \omega_{j,i} \varphi_P'\left(t; b_j, \alpha_{j,i}, \beta_{j,i}, \mu_{j,i}, \sigma_{j,i}\right), \quad j = 1, \dots, D,$$

and, for $t \neq 0$,

$$\begin{aligned}\varphi_P'\left(t; b, \alpha, \beta, \mu, \sigma\right) &= \frac{\partial}{\partial t}\varphi_P\left(t; b, \alpha, \beta, \mu, \sigma\right) \\ &= \varphi_P\left(t; b, \alpha, \beta, \mu, \sigma\right)\left(\frac{\alpha}{t}\nu + \imath\mu b\right)\end{aligned}$$

with

$$\nu = -\left|\sigma b t\right|^{\alpha}\left(1 - \imath\beta \operatorname{sgn}(bt) \tan\left(\alpha\frac{\pi}{2}\right)\right).$$

■

References

Adler, R. J. (1997). Discussion: Heavy Tail Modeling and Teletraffic Data. *The Annals of Statistics*, 25(5):1849–1852.

Alexander, C. and Lazar, E. (2006). Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling. *Journal of Applied Econometrics*, 21:307–336.

Alexander, C. and Lazar, E. (2009). Modelling Regime-Specific Stock Price Volatility. *Oxford Bulletin of Economics and Statistics*, 71(6):761–797.

Ausín, M. C. and Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, 51(5):2636–2652.

Badescu, A., Kulperger, R., and Lazar, E. (2008). Option Valuation with Normal Mixture GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, 12(2):5.

Bai, X., Russell, J. R., and Tiao, G. C. (2003). Kurtosis of GARCH and Stochastic Volatility Models with Non–Normal Innovations. *Journal of Econometrics*, 114:349–360.

Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2007). Multivariate Mixed Normal Conditional Heteroskedasticity. *Computational Statistics & Data Analysis*, 51(7):3551–3566.

Bauwens, L., Preminger, A., and Rombouts, J. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13(2):218–244.

Bauwens, L. and Storti, G. (2009). A Component GARCH Model with Time Varying Weights. *Studies in Nonlinear Dynamics & Econometrics*, 13(2):Article 1.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bonato, M. (2011). Modeling fat tails in stock returns: a multivariate stable-GARCH approach. *Computational Statistics*, 26:1–23.

Broda, S. A. (2011). Tail Probabilities and Partial Moments for Quadratic Forms in Multivariate Generalized Hyperbolic Random Vectors. Working Paper, University of Amsterdam.

Broda, S. A. and Paolella, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 7(4):412–436.

Broda, S. A. and Paolella, M. S. (2011). Expected Shortfall for Distributions in Finance. In Čížek, P., Härdle, W., and Rafał Weron, editors, *Statistical Tools for Finance and Insurance.*

Brooks, R. D., Faff, R. W., McKenzie, M. D., and Mitchell, H. (2000). A Multi-country Study of Power ARCH Models and National Stock Market Returns. *Journal of International Money and Finance*, 19(3):377–397.

Chen, Y., Härdle, W., and Spokoiny, V. (2006). GHICA — Risk Analysis with GH Distributions and Independent Components. SFB 649 Discussion Paper 2006-78, Humboldt-University, Berlin.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4):841–862.

Day, N. E. (1969). Estimating the Components of a Mixture of Normal Distributions. *Biometrika*, 56(3):463–474.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106.

Doganoglu, T., Hartz, C., and Mittnik, S. (2007). Portfolio Optimization when Risk Factors are Conditionally Varying and Heavy Tailed. *Computational Economics*, 29(3-4):333–354.

Doganoglu, T. and Mittnik, S. (1998). An Approximation Procedure for Asymmetric Stable Paretian Densities. *Computational Statistics*, 13:463–475.

Dowd, K. (2005). *Measuring Market Risk*. Wiley, New York, 2nd edition.

Fabricius, T., Kidmose, P., and Hansen, L. (2001). Dynamic components of linear stable mixtures from fractional low order moments. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 6:3957–3960.

Fama, E. F. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105.

Fan, J., Wang, M., and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:679–702.

Francq, C., Roussignol, M., and Zakoïan, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22(2):197–220.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Giacometti, R., Bertocchi, M. I., Rachev, S. T., and Fabozzi, F. J. (2007). Stable Distributions in the Black-Litterman Approach to Asset Allocation. *Quantitative Finance*, 7(4):423–433.

Gil-Pelaez, J. (1951). Note on the Inversion Theorem. *Biometrika*, 38:481–482.

Giot, P. and Laurent, S. (2003). Value-at-Risk for Long and Short Trading Positions. *Journal of Applied Econometrics*, 18(6):641–663.

Haas, M., Mittnik, S., and Paolella, M. S. (2004a). A New Approach to Markov Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.

Haas, M., Mittnik, S., and Paolella, M. S. (2009). Asymmetric Multivariate Normal Mixture GARCH. *Computational Statistics & Data Analysis*, 53(6):2129–2154.

Haas, M. and Paolella, M. S. (2011). Mixture and Regime-switching GARCH Models. In Bauwens, L., Hafner, C., and Laurent, S., editors, *Handbook of Volatility Models and Their Applications*. Wiley.

Hathaway, R. J. (1985). A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions. *The Annals of Statistics*, 13(2):795–800.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics*, 3(5):1163–1174.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley-Interscience, 1 edition.

Jansen, D. W. and de Vries, C. G. (1991). On the Frequency of Large Stock Returns: Putting Booms and Busts Into Perspective. *Review of Economics and Statistics*, 73:18–24.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62(1):49–66.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *The Annals of Mathematical Statistics*, 27(4):887–906.

Kiefer, N. M. (1978). Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model. *Econometrica*, 46(2):427–434.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2008). Financial Market Models with Lévy Processes and Time-Varying Volatility. *Journal of Banking & Finance*, 32(7):1363–1378.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2009a). Computing VaR and AVaR In Infinitely Divisible Distributions. Yale ICF Working Paper 09-07, Yale School of Management.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J. (2010). Tempered Stable and Tempered Infinitely Divisible GARCH Models. *Journal of Banking & Finance*, 34(9):2096–2109.

Kim, Y. S., Rachev, S. T., Chung, D. M., and Bianchi, M. L. (2009b). The Modified Tempered Stable Distribution, GARCH Models, and Option Pricing. *Probability and Mathematical Statistics*, 29:91–117.

Kirchler, M. and Huber, J. (2007). Fat Tails and Volatility Clustering in Experimental Asset Markets. *Journal of Economic Dynamics & Control*, 31:1844–1874.

Knight, J. and Satchell, S. (2001). *Return Distributions in Finance*. Butterworth-Heinemann, Oxford.

Kon, S. J. (1984). Models of Stock Returns: A Comparison. *The Journal of Finance*, 39(1):147–165.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press, Cambridge.

Kratz, M. and Resnick, S. I. (1996). The QQ-Estimator and Heavy Tails. *Communications in Statistics – Stochastic Models*, 12(4):699–724.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Lejeune, B. (2009). A Diagnostic m-test for Distributional Specification of Parametric Conditional Heteroscedasticity Models for Financial Data. *Journal of Empirical Finance*, 16(3):507–523.

Liu, J.-C. (2007). Stationarity for a Markov–switching Box–Cox Transformed Threshold GARCH Process. *Statistics and Probability Letters*, 77:1428–1438.

Lombardi, M. J. and Veredas, D. (2009). Indirect estimation of elliptical stable distributions. *Computational Statistics & Data Analysis*, 53(6):2309–2324.

Lyness, J. N. (1969). Notes on the adaptive Simpson quadrature routine. *Journal of the Association for Computing Machinery*, 16(3):483–495.

Malevergne, Y., Pisarenko, V., and Sornette, D. (2005). Empirical Distributions of Stock Returns: Between the Stretched Exponential and the Power Law? *Quantitative Finance*, 5:379–401.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business*, 36:394–419.

McCulloch, J. H. (1985). Interest-Risk Sensitive Deposit Insurance Premia: Stable ACH Estimates. *Journal of Banking and Finance*, 9:137–156.

McCulloch, J. H. (1996). Financial Applications of Stable Distributions. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics*, volume 14. Elsevier Science.

McCulloch, J. H. (1997). Measuring Tail Thickness in Order to Estimate the Stable Index α: A Critique. *Journal of Business and Economic Statistics*, 15:74–81.

McCulloch, J. H. (1998). Numerical Approximation of the Symmetric Stable Distribution and Density. In Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors, *A Practical Guide to Heavy Tails*, pages 489–499. Birkhäuser, Boston.

Mercuri, L. (2008). Option Pricing in a GARCH Model with Tempered Stable Innovations. *Finance Research Letters*, 5(3):172–182.

Mittnik, S., Doganoglu, T., and Chenyao, D. (1999). Computing the Probability Density Function of the Stable Paretian Distribution. *Mathematical and Computer Modelling*, 29:235–240.

Mittnik, S. and Paolella, M. S. (2003). Prediction of Financial Downside Risk with Heavy Tailed Conditional Distributions. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science, Amsterdam.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998). A Tail Estimator for the Index of the Stable Paretian Distribution. *Communications in Statistics–Theory and Methods*, 27(5):1239–1262.

Mittnik, S., Paolella, M. S., and Rachev, S. T. (2002). Stationarity of Stable Power–GARCH Processes. *Journal of Econometrics*, 106:97–107.

Nolan, J. P. (1997). Numerical Calculation of Stable Densities and Distribution Functions. *Stochastic Models*, 13(4):759–774.

Nolan, J. P. (1998). Univariate Stable Distributions: Parameterizations and Software. In Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors, *A Practical Guide to Heavy Tails*, pages 527–533. Birkhäuser, Boston.

Nolan, J. P. (2012). *Stable Distributions – Models for Heavy Tailed Data*. Birkhäuser, Boston. forthcoming; Chapter 1 online.

Paolella, M. S. (2001). Testing the Stable Paretian Assumption. *Mathematical and Computer Modelling*, 34:1095–1112.

Paolella, M. S. (2007). *Intermediate Probability: A Computational Approach*. John Wiley & Sons, Chichester.

Paolella, M. S. and Steude, S. C. (2008). Risk Prediction: A DWARF-like Approach. *Journal of Risk Model Validation*, 2(1):25–43.

Platen, E. and Heath, D. (2006). *A Benchmark Approach to Quantitative Finance*. Springer, Berlin, Germany.

Poirot, J. and Tankov, P. (2006). Monte Carlo Option Pricing for Tempered Stable (CGMY) Processes. *Asia-Pacific Financial Markets*, 13(4):327–344.

Rachev, S. T. (2003). *Handbook of Heavy Tailed Distributions in Finance*. North-Holland.

Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance.* Wiley & Sons, Chichester.

Rombouts, J. V. K. and Bouaddi, M. (2009). Mixed Exponential Power Asymmetric Conditional Heteroskedasticity. *Studies in Nonlinear Dynamics & Econometrics*, 13(3):Article 3.

Rombouts, J. V. K. and Stentoft, L. (2009). Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models. CREATES Research Papers 2009-07, School of Economics and Management, University of Aarhus.

Salas-Gonzalez, D., Kuruoglu, E. E., and Ruiz, D. P. (2009). Finite Mixtures of α-Stable Distributions. *Digital Signal Processing*, 19(2):250–264.

Samanidou, E., Zschischang, E., and Stauffer, D. (2007). Agent-Based Models of Financial Markets. *Reports on Progress in Physics*, 70:409–450.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman & Hall, London.

Seneta, E. (2004). Fitting the Variance-Gamma Model to Financial Data. *Journal of Applied Probability*, 41, Stochastic Methods and Their Applications:177–187.

Stoyanov, S., Samorodnitsky, G., Rachev, S., and Ortobelli, S. (2006). Computing the Portfolio Conditional Value-at-Risk in the alpha-stable Case. *Probability and Mathematical Statistics*, 26:1–22.

Sy, W. (2006). On the Coherence of VaR Risk Measure for Levy Stable Distributions. Working papers, Australian Prudential Regulation Authority.

Tanaka, K. (2009). Strong Consistency of the Maximum Likelihood Estimator for Finite Mixtures of LocationScale Distributions When Penalty is Imposed on the Ratios of the Scale Parameters. *Scandinavian Journal of Statistics*, 36:171–184.

Vigfusson, R. (1997). Switching between Chartists and Fundamentalists: A Markov Regime–Switching Approach. *International Journal of Finance and Economics*, 2:291–305.

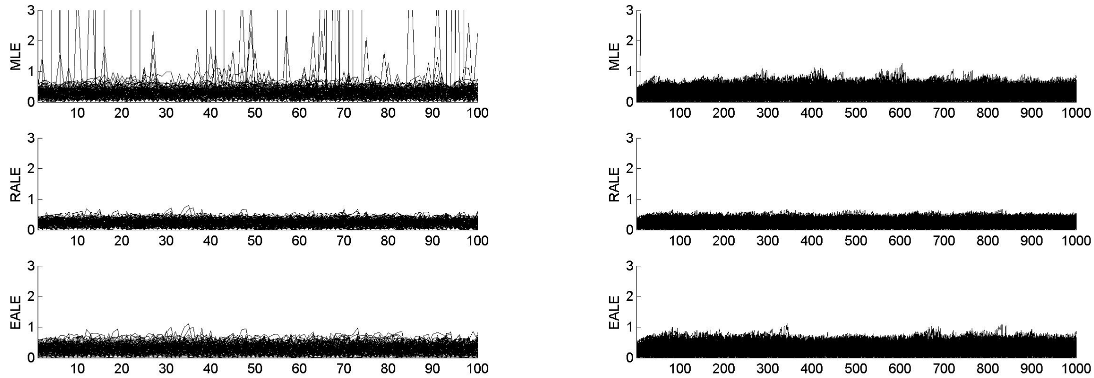
Vlaar, P. J. G. and Palm, F. C. (1993). The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps. *Journal of Business and Economic Statistics*, 11(3):351–360.

Weron, R. (2001). Levy-stable Distributions Revisited: Tail Index > 2 does not exclude the Levy-stable Regime. *International Journal of Modern Physics C*, 12:209–223.

Zolotarev, V. M. (1986). *One-Dimensional Stable Laws.* American Mathematical Society. Translation of Odnomernye Ustoichivye Raspredeleniia, NAUKA, Moskow, 1983.

MixStable(2, 2)

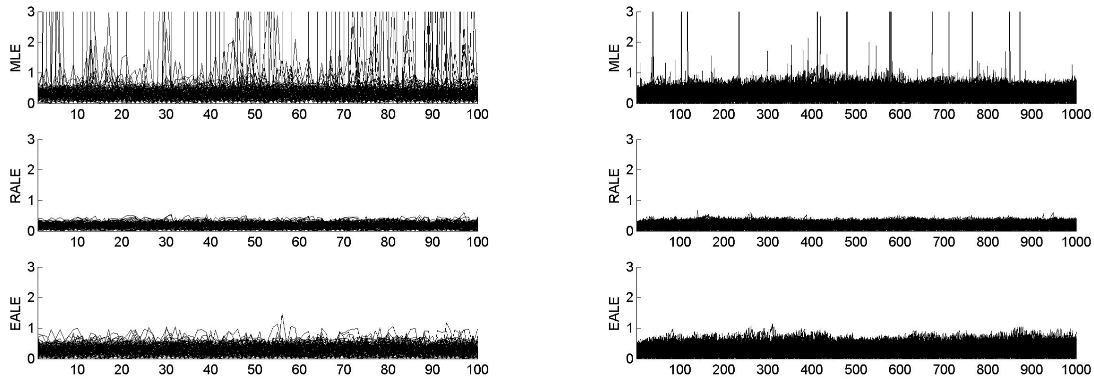
$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
0.315	4e-3	-	0.471	0.046	-	0.785	0.941	-	0.058	0.942	-	-0.721	0.044	-	2	0



sample size 100 | sample size 1000

MixStable(3, 3)

$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
1.012	4e-13	0.01	0.807	7e-3	0.089	0.648	0.984	0.916	0.019	0.346	0.635	-1.277	0.118	-0.025	2	0



sample size 100 | sample size 1000

Figure 2: Comparison of MLE, RALE and EALE in presence of local mixture degeneracy. Each panel shows 100 overlaid (maximized) likelihood series (not the augmented likelihood), from which the usual likelihood function is computed, obtained from estimates based on simulated data. Spikes depict either mixture degeneracy (singularities) or almost degenerated local optima. Results are essentially identical for all mixture GARCH models under study. The data generating process roughly corresponds to estimates found in our empirical analysis of DJIA return data.

MixStable$(2,2)$

$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
0.004	0.04	-	0.02	0.09	-	0.97	0.91	-	0.07	0.93	-	0.67	-0.05	-	1.96	0

estimator	sample size	P-RMSE arith.	P-RMSE log-geo.		Q-RMSE arith.	Q-RMSE log-geo.	A-RMSE arith.	A-RMSE log-geo.		P-IQR arith.	P-IQR log-geo.	Q-IQR arith.	Q-IQR log-geo.	A-IQR arith.	A-IQR log-geo.
MLE	100	3.56	-60.32	(512)	2.72	-1.70	0.65	-5.98	(84)	3.73	-60.55	1.71	-1.99	0.20	-6.55
RALE		3.70	**-60.20**	(179)	**1.92**	**-2.04**	**0.64**	**-6.00**	(1)	**3.86**	**-60.20**	1.66	**-2.08**	0.23	-6.54
EALE		**3.68**	-60.19	(291)	2.57	-1.80	0.64	-6.00	(106)	4.31	-60.12	**1.65**	-2.03	**0.21**	**-6.56**
MLE	1000	2.47	-60.89	(270)	2.07	-2.11	0.56	-6.03	(17)	2.46	-61.05	0.95	-2.68	0.26	-6.33
RALE		2.93	**-60.57**	(65)	**1.65**	**-2.14**	**0.55**	**-6.05**	(0)	**2.35**	-60.76	**1.04**	**-2.61**	**0.24**	**-6.34**
EALE		**2.67**	-60.65	(106)	1.71	-2.12	0.56	-6.03	(11)	2.55	**-60.78**	1.06	-2.53	0.25	-6.33
MLE	10000	2.66	-61.15	(222)	1.72	-2.24	0.53	-6.07	(14)	2.26	-61.42	0.41	-3.44	0.18	-6.48
RALE		**1.94**	**-61.06**	(25)	**1.67**	**-2.23**	**0.51**	**-6.07**	(0)	**1.83**	-61.28	**0.40**	**-3.48**	0.18	-6.47
EALE		2.36	-61.03	(65)	1.96	-1.91	0.53	-6.06	(1)	1.89	**-61.43**	0.42	-3.43	**0.18**	**-6.47**

MixStable$(3,3)$

$\gamma_{0,1}$	$\gamma_{0,2}$	$\gamma_{0,3}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{3,1}$	Ψ_1	Ψ_2	Ψ_3	ω_1	ω_2	ω_3	μ_1	μ_2	μ_3	α	β
0.4	0.005	0.07	0.41	0.01	0.09	0.71	0.98	0.92	0.03	0.34	0.63	1.28	0.12	-0.125	1.93	0

estimator	sample size	P-RMSE arith.	P-RMSE log-geo.		Q-RMSE arith.	Q-RMSE log-geo.	A-RMSE arith.	A-RMSE log-geo.		P-IQR arith.	P-IQR log-geo.	Q-IQR arith.	Q-IQR log-geo.	A-IQR arith.	A-IQR log-geo.
MLE	100	7.00	-42.70	(781)	7.02	-0.89	6.55	-2.73	(58)	7.79	-43.82	3.82	-1.42	0.15	-6.99
RALE		6.57	-42.72	(457)	9.06	-1.01	0.54	-5.95	(32)	**6.64**	**-42.68**	**3.54**	-1.44	0.20	-6.83
EALE		**6.55**	**-42.74**	(416)	**8.54**	**-1.04**	**0.53**	**-5.97**	(125)	7.01	-42.67	3.63	**-1.44**	**0.19**	**-6.89**
MLE	1000	5.99	-42.93	(473)	2.66	-1.93	0.46	-5.99	(13)	7.86	-43.15	2.17	-2.12	0.16	-6.76
RALE		5.60	-42.95	(182)	**2.31**	**-1.99**	**0.44**	**-6.02**	(0)	**4.91**	-43.04	2.27	-1.93	0.20	-6.51
EALE		**5.45**	**-42.97**	(91)	2.71	-1.83	0.46	-5.99	(14)	5.04	**-43.11**	**2.04**	**-2.08**	**0.17**	**-6.62**
MLE	10000	4.37	-43.31	(401)	2.17	-2.10	0.44	-6.03	(6)	26.59	-43.69	1.28	-2.57	0.15	-6.73
RALE		4.40	-43.21	(103)	2.42	-1.93	**0.41**	**-6.05**	(0)	3.81	-43.49	1.18	-2.60	0.16	-6.59
EALE		**4.10**	**-43.30**	(59)	**2.16**	**-2.09**	0.44	-6.03	(2)	**3.38**	**-43.90**	**1.14**	**-2.61**	**0.15**	**-6.66**

Table 1: Aggregated goodness-of-fit measures based on either parameter errors (P), quantile errors (Q) or acf errors (A) computed from $N = 1000$ estimates based on simulated data. Given values represent error sums, either $\sum_j \xi\left(\epsilon^{\tau}_{j,\cdot}\right)$ (arithmetic) or $1/N \sum_j \log \xi\left(\epsilon^{\tau}_{j,\cdot}\right)$ (log-geometric), where $\xi \in \{\mathrm{RMSE}, \mathrm{IQR}\}$ and $\tau \in \{\mathrm{P}, \mathrm{Q}, \mathrm{A}\}$. As the geometric mean depicts the common factor of the product (being a lower bound to the arithmetic mean), it gives a more robust measure of the aggregated error less affected by outliers than its arithmetic counterpart; without loss of generality we report its log value for reasons of scaling. Values in parentheses denote the number of removed estimates as described in Section 3.3. Similar to the case of the P-RMSE, estimates are removed for Q- and A-RMSE where $\sigma_{i,t} \to 0$ or $\sigma_{i,t} \to \infty$. The standard maximum likelihood estimator (gray) is excluded as its likelihood function is ill-defined for mixtures. Entries in boldface denote the smallest error.

data set	model	GARCH power coefficient δ										
		1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
DAX	MixNormal(4, 4)	-8091.78	-8090.64	-8090.11	-8090.15	-8090.70	-8091.71	**-8093.13**	-8094.89	-8096.94	-8099.25	-8101.68
	MixGED(4, 4)	-8091.78	-8090.64	**-8090.11**	**-8090.15**	**-8090.70**	**-8091.71**	-8093.13	**-8094.89**	-8096.94	-8099.25	-8101.68
	A^1MixStable(4, 4)	**-8091.66**	**-8090.58**	-8090.12	-8090.15	-8090.71	-8091.72	-8093.13	-8094.89	**-8096.93**	**-8099.12**	**-8101.51**
	A^2MixStable(4, 4)	-8103.16	-8102.01	-8101.42	-8101.41	-8101.89	-8102.80	-8104.10	-8105.71	-8107.61	-8109.76	-8112.11
S&P	MixNormal(4, 4)	-6678.03	-6676.27	-6674.94	-6673.98	-6673.37	-6673.07	-6673.06	-6673.31	-6673.80	-6674.51	-6675.42
	MixGED(4, 4)	-6678.03	-6676.27	-6674.94	-6673.98	-6673.37	-6673.07	-6673.06	-6673.31	-6673.80	-6674.51	-6675.42
	A^1MixStable(4, 4)	**-6677.89**	**-6676.13**	**-6674.80**	**-6673.85**	**-6673.24**	**-6672.95**	**-6672.94**	**-6673.20**	**-6673.69**	**-6674.40**	**-6675.32**
	A^2MixStable(4, 4)	-6686.20	-6684.35	-6683.03	-6681.95	-6681.23	-6680.82	-6680.69	-6680.83	-6681.20	-6681.79	-6682.57
DJIA	MixNormal(4, 4)	-6632.67	-6631.47	-6630.62	-6630.10	-6629.87	-6629.92	-6630.21	-6630.73	-6631.45	-6632.35	-6633.42
	MixGED(4, 4)	-6632.72	-6631.50	-6630.62	-6630.10	-6629.87	-6629.92	-6630.21	-6630.73	-6631.45	-6632.35	-6633.42
	A^1MixStable(4, 4)	**-6632.26**	**-6631.08**	**-6630.24**	**-6629.63**	**-6629.33**	**-6629.29**	**-6629.49**	**-6629.93**	**-6630.56**	**-6631.37**	**-6632.36**
	A^2MixStable(4, 4)	-6639.28	-6637.99	-6637.04	-6636.41	-6636.06	-6635.97	-6636.12	-6636.49	-6637.06	-6637.81	-6638.72
NIKKEI	MixNormal(4, 4)	-8459.40	-8458.30	-8457.62	-8457.31	-8457.33	-8457.66	-8458.27	-8459.13	-8460.20	-8461.48	-8462.95
	MixGED(4, 4)	-8459.40	-8458.30	-8457.62	-8457.31	-8457.33	-8457.66	-8458.27	-8459.13	-8460.20	-8461.48	-8462.95
	A^1MixStable(4, 4)	**-8458.99**	**-8457.90**	**-8457.23**	**-8456.93**	**-8456.94**	**-8457.26**	**-8457.85**	**-8458.68**	**-8459.73**	**-8460.97**	**-8462.38**
	A^2MixStable(4, 4)	-8464.85	-8463.98	-8463.55	-8463.51	-8463.80	-8464.39	-8465.20	-8466.26	-8467.51	-8468.95	-8470.54
¥/€	MixNormal(4, 4)	-2627.83	-2626.89	-2626.16	-2625.60	-2625.19	-2624.91	-2624.76	-2624.70	-2624.75	-2624.87	-2625.07
	MixGED(4, 4)	-2627.83	-2626.89	-2626.16	-2625.60	-2625.19	-2624.91	-2624.76	-2624.70	-2624.75	-2624.87	-2625.07
	A^1MixStable(4, 4)	**-2627.83**	**-2626.89**	**-2626.15**	**-2625.59**	**-2625.19**	**-2624.91**	**-2624.76**	**-2624.70**	**-2624.74**	**-2624.87**	**-2625.07**
	A^2MixStable(4, 4)	-2638.27	-2637.16	-2636.22	-2635.49	-2634.92	-2634.51	-2634.22	-2634.05	-2633.98	-2634.01	-2634.12
$/€	MixNormal(4, 4)	-2447.20	-2445.99	-2444.99	-2444.17	-2443.51	-2443.01	-2442.63	-2442.37	-2442.22	-2442.16	-2442.18
	MixGED(4, 4)	-2445.46	-2444.56	-2443.81	-2443.20	-2442.71	-2442.33	-2442.06	-2441.88	-2441.78	-2441.76	-2441.82
	A^1MixStable(4, 4)	**-2443.89**	**-2443.23**	**-2442.74**	**-2442.40**	**-2442.18**	**-2442.07**	**-2441.90**	**-2441.71**	**-2441.59**	**-2441.52**	**-2441.50**
	A^2MixStable(4, 4)	-2447.76	-2447.09	-2446.56	-2446.15	-2445.84	-2445.65	-2445.42	-2445.27	-2445.18	-2445.17	-2445.22
NASDAQ	MixNormal(4, 4)	-8138.21	-8137.30	-8136.79	-8136.62	-8136.77	-8137.19	-8137.86	-8138.75	-8139.85	-8141.13	-8142.57
	MixGED(4, 4)	-8138.21	-8137.30	-8136.79	-8136.62	-8136.77	-8137.19	-8137.86	-8138.75	-8139.85	-8141.13	-8142.57
	A^1MixStable(4, 4)	**-8134.43**	**-8133.49**	**-8132.96**	**-8132.78**	**-8132.90**	**-8133.30**	**-8133.95**	**-8134.82**	**-8135.88**	**-8137.12**	**-8138.53**
	A^2MixStable(4, 4)	-8139.44	-8138.49	-8137.92	-8137.71	-8137.80	-8138.17	-8138.79	-8139.63	-8140.67	-8141.89	-8143.25

Table 2: Profile likelihood results for different GARCH power parameters δ and all 4-component models and data sets under study. Each row corresponds to fitting the model to the entire return series of 20 years (equity) or 10 years (FX), respectively. Entries in boldface denote the best (in-sample) fits. Underlined numbers denote the best outcome per model. Gray values denote MixStable-GARCH estimates for which the power restriction $\delta < \alpha$ (necessary for $\mathbb{E}[|\varepsilon_t|^\delta] < \infty$) is violated.

model	δ	DAX	S&P	DJIA	NIKKEI	¥/€	\$/€	NASDAQ
Anderson-Darling								
MixNormal(2, 2)	1	0.86	1.00	1.27	**0.93**	**1.54**	**1.52**	**1.32**
	2	**0.71**	**0.89**	**1.19**	1.10	1.72	2.01*	1.37
MixNormal(3, 2)	1	0.65	**0.54**	**0.59**	**0.73**	**1.49**	**1.33**	**0.43**
	2	**0.57**	0.67	0.59	0.76	1.53	1.92	0.49
MixNormal(3, 3)	1	0.55	0.53	**0.54**	**0.60**	**1.41**	**1.51**	**0.42**
	2	**0.52**	**0.50**	0.63	0.70	1.47	1.99*	0.43
MixNormal(4, 4)	1	0.69	**0.37**	**0.42**	**0.54**	**1.22**	**1.59**	**0.47**
	2	**0.58**	0.42	0.66	0.63	1.29	1.61	0.48
MixGED(2, 2)	1	0.72	**0.63**	**0.67**	**0.73**	**1.44**	**1.41**	**0.30**
	2	**0.64**	0.71	0.68	0.85	1.72	1.83	0.59
MixGED(3, 2)	1	0.76	0.56	0.52	**0.62**	**1.45**	**1.09**	0.52
	2	**0.73**	**0.53**	**0.46**	0.73	1.59	1.81	**0.33**
MixGED(3, 3)	1	0.68	**0.51**	**0.49**	**0.48**	1.45	**1.30**	0.46
	2	**0.56**	0.51	0.82	0.71	**1.45**	1.87	**0.38**
MixGED(4, 4)	1	0.72	**0.41**	**0.59**	**0.45**	**1.21**	**1.47**	0.51
	2	**0.58**	0.56	0.82	0.65	1.23	1.79	**0.46**
Cramér-von Mises								
MixNormal(2, 2)	1	0.12	0.12	0.22	**0.16**	**0.30**	**0.30**	0.23
	2	**0.11**	**0.11**	**0.20**	0.18	0.34	0.41*	**0.19**
MixNormal(3, 2)	1	**0.06**	**0.07**	0.07	0.12	**0.27**	**0.25**	**0.04**
	2	0.07	0.08	**0.06**	**0.11**	0.28	0.39*	0.06
MixNormal(3, 3)	1	0.06	0.08	**0.06**	0.08	**0.25**	**0.28**	**0.04**
	2	**0.05**	**0.06**	0.07	**0.07**	0.28	0.39*	0.04
MixNormal(4, 4)	1	0.07	**0.06**	**0.05**	0.08	**0.22**	**0.29**	0.04
	2	**0.06**	0.06	0.07	**0.07**	0.23	0.30	**0.04**
MixGED(2, 2)	1	**0.09**	**0.08**	**0.09**	**0.10**	**0.27**	**0.26**	**0.03**
	2	0.09	0.08	0.10	0.13	0.34	0.36*	0.06
MixGED(3, 2)	1	**0.08**	0.07	0.06	**0.09**	**0.26**	**0.20**	0.05
	2	0.09	**0.06**	**0.05**	0.10	0.29	0.36*	**0.03**
MixGED(3, 3)	1	0.07	0.07	**0.06**	**0.06**	**0.26**	**0.24**	0.04
	2	**0.06**	**0.06**	0.09	0.07	0.27	0.36*	**0.04**
MixGED(4, 4)	1	0.07	**0.05**	**0.07**	**0.06**	0.22	**0.27**	0.05
	2	**0.06**	0.07	0.09	0.06	**0.22**	0.33	**0.04**
Integrated root mean squared error, IRMSE, up to the 5% VaR level								
MixNormal(2, 2)	1	**0.19**	0.36	**0.22**	**0.31**	**0.34**	0.53	**0.20**
	2	0.21	**0.32**	0.29	0.39	0.36	**0.51**	0.25
MixNormal(3, 2)	1	**0.17**	**0.25**	**0.16**	**0.28**	**0.47**	0.69	0.49
	2	0.22	0.27	0.16	0.35	0.51	**0.49**	**0.35**
MixNormal(3, 3)	1	**0.10**	0.23	**0.10**	**0.28**	0.47	0.62	0.53
	2	0.22	**0.22**	0.14	0.34	**0.46**	**0.37**	**0.37**
MixNormal(4, 4)	1	**0.20**	0.22	**0.10**	**0.27**	**0.42**	0.59	0.64
	2	0.24	**0.19**	0.11	0.37	0.48	**0.40**	**0.51**
MixGED(2, 2)	1	**0.20**	**0.27**	**0.16**	**0.31**	0.40	**0.31**	0.35
	2	0.23	0.28	0.25	0.34	**0.36**	0.34	**0.23**
MixGED(3, 2)	1	**0.19**	0.29	**0.12**	**0.26**	**0.46**	0.54	0.60
	2	0.25	**0.23**	0.13	0.35	0.50	**0.34**	**0.32**
MixGED(3, 3)	1	**0.16**	**0.23**	**0.12**	**0.28**	0.47	**0.33**	0.61
	2	0.24	0.26	0.15	0.34	**0.44**	0.39	**0.39**
MixGED(4, 4)	1	**0.17**	**0.22**	**0.10**	**0.27**	**0.44**	**0.37**	0.64
	2	0.25	0.24	0.11	0.38	0.47	0.45	**0.54**

Table 3: Anderson-Darling and Cramér-von Mises test statistics as well as IRMSE values for all MixNormal- and MixGED-GARCH models and data sets under study. Entries in boldface denote the best outcomes. For AD and CM, ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. All values are based on evaluating the one-step-ahead out-of-sample distribution forecasts at the observed return data, estimated throughout a rolling window of 1,000 trading days (updated every 20 days), covering 10 years of equity returns (July 7, 1999 to July 7, 2009) and 5 years (July 7, 2004 to July 7, 2009) of FX returns.

	A^1Mix-Stable $(2,2)$	A^1Mix-Stable $(3,2)$	A^1Mix-Stable $(3,3)$	A^1Mix-Stable $(4,4)$	A^2Mix-Stable $(2,2)$	A^2Mix-Stable $(3,2)$	A^2Mix-Stable $(3,3)$	A^2Mix-Stable $(4,4)$
$\gamma_{0,1}$	4.8e-13	1.6e-20	0.130	1.2e-20	2.1e-18	3.4e-13	7.1e-14	0.003
	(1.3e-15)	(9.5e-27)	(0.049)	(1.4e-27)	(9.9e-24)	(5.7e-16)	(9.6e-17)	(0.001)
$\gamma_{0,2}$	0.012	0.012	8.9e-25	0.010	0.007	0.007	0.008	2.7e-20
	(0.004)	(0.003)	(1.8e-39)	(0.003)	(0.002)	(0.002)	(0.003)	(8.6e-27)
$\gamma_{0,3}$		2.682	0.005	0.438		0.964	0.002	0.002
		(0.387)	(0.002)	(0.143)		(3.028)	(0.001)	(0.001)
$\gamma_{0,4}$				6.0e-13				0.042
				(5.2e-16)				(0.016)
$\gamma_{1,1}$	0.015	0.014	0.339	0.014	0.010	0.010	0.011	0.119
	(0.004)	(0.003)	(0.090)	(0.004)	(0.003)	(0.002)	(0.003)	(0.018)
$\gamma_{1,2}$	0.112	0.105	0.011	0.123	0.074	0.074	0.118	0.010
	(0.013)	(0.013)	(0.003)	(0.018)	(0.007)	(0.008)	(0.016)	(0.003)
$\gamma_{1,3}$			0.067	0.442			0.021	0.021
			(0.008)	(0.140)			(0.005)	(0.005)
$\gamma_{1,4}$				0.023				0.101
				(0.005)				(0.042)
ψ_1	0.984	0.984	0.778	0.970	0.981	0.981	0.976	0.911
	(0.004)	(0.003)	(0.057)	(0.008)	(0.006)	(0.005)	(0.007)	(0.013)
ψ_2	0.911	0.917	0.978	0.904	0.936	0.936	0.908	0.976
	(0.010)	(0.009)	(0.006)	(0.013)	(0.006)	(0.006)	(0.011)	(0.008)
ψ_3			0.943	0.663			0.979	0.979
			(0.007)	(0.080)			(0.005)	(0.005)
ψ_4				0.980				0.893
				(0.004)				(0.033)
ω_1	0.439	0.420	0.046	0.101	0.145	0.145	0.089	0.438
	(0.024)	(0.057)	(0.012)	(0.024)	(0.032)	(0.037)	(0.020)	(0.065)
ω_2	0.561	0.571	0.131	0.477	0.855	0.855	0.498	0.086
	(red.)	(0.007)	(0.039)	(0.061)	(red.)	(4.1e-6)	(0.066)	(0.033)
ω_3		0.009	0.823	0.019		1.6e-10	0.413	0.408
		(red.)	(red.)	(0.011)		(red.)	(red.)	(0.107)
ω_4				0.403				0.068
				(red.)				(red.)
μ_1	0.138	0.118	0.846	0.048				
	(0.033)	(0.026)	(0.165)	(0.039)				
μ_2	0.108	0.054	0.081	0.129				
	(red.)	(0.030)	(0.037)	(0.055)				
μ_3		2.028	0.035	1.226				
		(red.)	(red.)	(0.195)				
μ_4				0.197				
				(red.)				
α	1.917	2.000	1.989	1.997	1.890	1.890	1.892	1.894
	(0.025)	(0.004)	(4.7e-5)	(0.010)	(0.013)	(0.013)	(0.013)	(0.013)
β					0.707	0.706	0.744	0.781
					(0.061)	(0.065)	(0.027)	(0.059)
LL	-6652.061	-6641.523	-6639.499	-6632.264	-6644.757	-6644.756	-6639.442	-6639.277

Table 4: Maximum likelihood parameter estimates for the MixStable-GARCH models under study based on the entire set of DJIA return data (20 years). Standard errors are given in parentheses. The shortcut "red." stands for "redundant" and indicates that the parameter is not estimated but recovered from the corresponding imposed constraints.

model	free param.	crit.	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
MixNormal(2, 2)	8	LL	-8120.9	-6716.4	-6660.3	-8476.0	-2634.8	-2455.1	-8163.5
		AIC	16257.8	13448.8	13336.6	16968.0	**5285.5**	4926.1	16343.1
		BIC	16310.0	13501.0	13388.7	17020.0	**5332.2**	4972.8	16395.4
MixNormal(3, 2)	11	LL	-8097.2	-6688.3	-6641.5	-8466.7	-2632.7	-2450.2	-8151.8
		AIC	**16216.3**	13398.6	13305.0	16955.4	5287.3	4922.4	16325.5
		BIC	16288.1	13470.4	13376.8	17027.0	5351.5	4986.6	16397.4
MixNormal(3, 3)	13	LL	-8097.1	-6686.1	-6639.3	-8465.2	-2629.8	-2449.6	-8146.1
		AIC	16220.3	13398.3	13304.7	16956.4	5285.6	4925.3	16318.2
		BIC	16305.1	13483.1	13389.5	17041.0	5361.4	5001.1	16403.1
MixNormal(4, 4)	18	LL	-8091.8	-6678.0	-6632.7	-8459.4	-2627.8	-2447.2	-8138.2
		AIC	16219.6	**13392.1**	13301.3	16954.8	5291.7	4930.4	16312.4
		BIC	16337.0	13509.5	13418.7	17071.8	5396.6	5035.4	16430.0
MixGED(2, 2)	9	LL	-8116.2	-6693.3	-6644.9	-8471.1	-2634.8	-2446.3	-8158.6
		AIC	16250.4	13404.6	13307.9	16960.1	5287.5	**4910.5**	16335.2
		BIC	16309.2	13463.3	13366.6	**17018.6**	5340.0	**4963.1**	16394.0
MixGED(3, 2)	12	LL	-8097.0	-6686.7	-6637.6	-8465.1	-2632.6	-2445.7	-8151.1
		AIC	16218.0	13397.4	**13299.2**	**16954.3**	5289.3	4915.5	16326.2
		BIC	16296.3	13475.7	13377.5	17032.3	5359.2	4985.5	16404.6
MixGED(3, 3)	14	LL	-8097.0	-6685.5	-6636.7	-8464.4	-2629.8	-2445.5	-8146.1
		AIC	16222.0	13399.1	13301.4	16956.7	5287.6	4918.9	16320.2
		BIC	16313.3	13490.4	13392.7	17047.7	5369.2	5000.6	16411.6
MixGED(4, 4)	19	LL	-8091.8	-6678.0	-6632.7	-8459.4	-2627.8	-2445.5	-8138.2
		AIC	16221.6	13394.1	13303.4	16956.8	5293.7	4928.9	16314.4
		BIC	16345.5	13518.0	13427.4	17080.3	5404.4	5039.8	16438.6
A^{1}MixStable(2, 2)	9	LL	-8103.0	-6699.9	-6652.1	-8471.2	-2634.4	-2450.1	-8159.8
		AIC	16224.0	13417.8	13322.1	16960.4	5286.8	4918.1	16337.7
		BIC	**16282.7**	13476.6	13380.8	17018.9	5339.3	4970.6	16396.5
A^{1}MixStable(3, 2)	12	LL	-8097.2	-6688.3	-6641.5	-8466.7	-2632.7	-2447.7	-8137.4
		AIC	16218.3	13400.7	13307.0	16957.4	5289.3	4919.3	**16298.8**
		BIC	16296.6	13478.9	13385.3	17035.5	5359.3	4989.4	16377.3
A^{1}MixStable(3, 3)	14	LL	-8096.5	-6686.1	-6639.5	-8463.8	-2629.8	-2446.3	-8142.2
		AIC	16221.0	13400.3	13307.0	16955.6	5287.6	4920.7	16312.4
		BIC	16312.4	13491.6	13398.3	17046.7	5369.2	5002.4	16403.9
A^{1}MixStable(4, 4)	19	LL	**-8091.7**	**-6677.9**	**-6632.3**	**-8459.0**	**-2627.8**	**-2443.9**	**-8134.4**
		AIC	16221.3	13393.8	13302.5	16956.0	5293.7	4925.8	16306.9
		BIC	16345.3	13517.7	13426.5	17079.5	5404.4	5036.6	16431.0
A^{2}MixStable(2, 2)	9	LL	-8103.3	-6692.0	-6644.8	-8475.0	-2638.5	-2450.7	-8146.9
		AIC	16224.6	13402.0	13307.5	16968.0	5295.1	4919.5	16311.7
		BIC	16283.3	**13460.7**	**13366.2**	17026.5	5347.5	4972.0	**16370.6**
A^{2}MixStable(3, 2)	12	LL	-8103.3	-6692.0	-6644.8	-8471.0	-2638.5	-2450.0	-8147.7
		AIC	16228.6	13406.0	13311.5	16964.1	5299.1	4921.9	16317.4
		BIC	16300.4	13477.8	13383.3	17035.6	5363.2	4986.1	16389.3
A^{2}MixStable(3, 3)	13	LL	-8103.2	-6686.3	-6639.4	-8468.1	-2638.4	-2448.8	-8141.1
		AIC	16232.3	13398.5	13304.9	16962.2	5302.9	4923.6	16308.2
		BIC	16317.2	13483.3	13389.7	17046.7	5378.7	4999.4	16393.2
A^{2}MixStable(4, 4)	17	LL	-8103.2	-6686.2	-6639.3	-8464.9	-2638.3	-2447.8	-8139.4
		AIC	16240.3	13406.4	13312.6	16963.7	5310.5	4929.5	16312.9
		BIC	16351.3	13517.3	13423.4	17074.2	5409.6	5028.7	16424.0

Table 5: Estimation results for all multi-component mixture GARCH models using $\delta = 1$ and all data sets under study (20 years of equity returns, 10 years of FX returns). The three-row blocks show the log-likelihood value (LL) and the corresponding information criteria (AIC and BIC). Entries in boldface denote the best column-wise (in-sample) fits.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Anderson-Darling							
MixNormal(2, 2)	0.86	1.00	1.27	0.93	1.54	1.52	1.32
MixNormal(3, 2)	0.65	0.54	0.59	0.73	1.49	1.33	0.43
MixNormal(3, 3)	**0.55**	0.53	0.54	0.60	1.41	1.51	0.42
MixNormal(4, 4)	0.69	0.37	0.42	0.54	1.22	1.59	0.47
MixGED(2, 2)	0.72	0.63	0.67	0.73	1.44	1.41	**0.30**
MixGED(3, 2)	0.76	0.56	0.52	0.62	1.45	1.09	0.52
MixGED(3, 3)	0.68	0.51	0.49	0.48	1.45	1.30	0.46
MixGED(4, 4)	0.72	0.41	0.59	0.45	1.21	1.47	0.51
A^1MixStable(2, 2)	0.73	0.49	0.55	0.92	1.36	1.20	0.35
A^1MixStable(3, 2)	0.70	0.45	0.43	0.65	1.48	**1.01**	0.31
A^1MixStable(3, 3)	0.66	0.43	0.36	0.44	1.43	1.38	0.36
A^1MixStable(4, 4)	0.63	**0.31**	**0.28**	**0.34**	1.24	1.48	0.44
A^2MixStable(2, 2)	1.12	1.05	0.67	0.59	**0.84**	1.88	1.61
A^2MixStable(3, 2)	1.16	1.02	0.58	0.58	0.85	1.71	1.59
A^2MixStable(3, 3)	1.22	0.95	0.56	0.58	0.85	1.68	1.59
A^2MixStable(4, 4)	1.23	0.99	0.56	0.59	0.86	1.70	1.63
Cramér-von Mises							
MixNormal(2, 2)	0.12	0.12	0.22	0.16	0.30	0.30	0.23
MixNormal(3, 2)	0.06	0.07	0.07	0.12	0.27	0.25	0.04
MixNormal(3, 3)	**0.06**	0.08	0.06	0.08	0.25	0.28	0.04
MixNormal(4, 4)	0.07	0.06	0.05	0.08	0.22	0.29	0.04
MixGED(2, 2)	0.09	0.08	0.09	0.10	0.27	0.26	**0.03**
MixGED(3, 2)	0.08	0.07	0.06	0.09	0.26	0.20	0.05
MixGED(3, 3)	0.07	0.07	0.06	0.06	0.26	0.24	0.04
MixGED(4, 4)	0.07	0.05	0.07	0.06	0.22	0.27	0.05
A^1MixStable(2, 2)	0.09	0.07	0.08	0.15	0.25	0.23	0.04
A^1MixStable(3, 2)	0.08	0.05	0.05	0.10	0.27	**0.19**	0.03
A^1MixStable(3, 3)	0.06	0.06	0.04	0.05	0.26	0.24	0.04
A^1MixStable(4, 4)	0.06	**0.05**	**0.03**	**0.04**	0.22	0.26	0.04
A^2MixStable(2, 2)	0.14	0.15	0.10	0.07	**0.15**	0.41*	0.27
A^2MixStable(3, 2)	0.14	0.14	0.08	0.06	0.16	0.37*	0.26
A^2MixStable(3, 3)	0.15	0.13	0.07	0.06	0.16	0.36*	0.27
A^2MixStable(4, 4)	0.15	0.14	0.07	0.06	0.16	0.37*	0.27
Kolmogorov-Smirnov (test statistics are scaled up by factor 100)							
MixNormal(2, 2)	1.48	1.44	1.97	1.73	3.23	3.76*	2.41
MixNormal(3, 2)	1.28	1.64	1.36	1.47	3.39	3.47	1.07
MixNormal(3, 3)	**1.19**	1.57	1.25	1.25	3.40	3.58	0.96*
MixNormal(4, 4)	1.38	1.32	1.17	1.34	3.12	3.71	1.19
MixGED(2, 2)	1.33	1.62	1.33	1.44	3.13	3.63	1.09
MixGED(3, 2)	1.36	1.47	1.21	1.26	3.39	**2.77**	1.05
MixGED(3, 3)	1.33	1.50	1.36	1.15	3.40	3.33	0.97*
MixGED(4, 4)	1.42	**1.06**	1.34	1.03	3.19	3.37	1.06
A^1MixStable(2, 2)	1.29	1.25	1.31	1.73	3.11	3.14	1.11
A^1MixStable(3, 2)	1.27	1.54	1.21	1.59	3.32	2.79	**0.86****
A^1MixStable(3, 3)	1.33	1.48	1.21	1.10	3.40	3.33	1.04
A^1MixStable(4, 4)	1.28	1.11	**1.02**	**0.90****	3.25	3.35	1.14
A^2MixStable(2, 2)	2.17	1.63	1.74	1.54	**2.86**	4.27**	2.18
A^2MixStable(3, 2)	2.21	1.79	1.30	1.41	2.87	3.98*	2.03
A^2MixStable(3, 3)	2.21	1.67	1.26	1.46	2.89	3.88*	2.16
A^2MixStable(4, 4)	2.20	1.77	1.32	1.51	2.89	3.88*	2.16

Table 6: Anderson-Darling, Cramér-von Mises and Kolmogorov-Smirnov test statistics for all models (with $\delta = 1$) and data sets under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
Ljung-Box ($m = 20$ lags)							
MixNormal$(2, 2)$	26.44	32.97*	30.66	16.95	17.46	18.57	31.05
MixNormal$(3, 2)$	26.74	33.15*	30.46	17.01	17.59	18.55	31.62*
MixNormal$(3, 3)$	26.63	33.26*	30.68	17.49	17.65	18.50	30.92
MixNormal$(4, 4)$	**26.25**	33.77*	30.78	17.95	17.99	18.03	31.86*
MixGED$(2, 2)$	26.62	33.61*	31.12	**16.94**	17.46	**17.55**	30.92
MixGED$(3, 2)$	26.73	33.15*	30.69	17.12	17.78	17.95	31.83*
MixGED$(3, 3)$	26.90	33.41*	31.33	17.96	17.64	17.96	31.37
MixGED$(4, 4)$	26.51	33.61*	30.62	18.10	17.82	18.20	31.46*
A^{1}MixStable$(2, 2)$	26.46	**32.87***	**29.83**	16.96	17.34	18.49	**30.49**
A^{1}MixStable$(3, 2)$	26.29	33.35*	30.63	19.64	17.67	19.17	31.50*
A^{1}MixStable$(3, 3)$	26.33	33.55*	30.72	17.59	17.77	17.94	30.78
A^{1}MixStable$(4, 4)$	26.44	33.94*	31.25	18.24	17.67	18.29	31.45*
A^{2}MixStable$(2, 2)$	27.21	33.45*	30.00	17.83	16.17	17.66	31.30
A^{2}MixStable$(3, 2)$	26.95	33.12*	29.86	17.92	**16.14**	17.74	31.58*
A^{2}MixStable$(3, 3)$	27.00	33.14*	29.99	17.89	16.22	17.94	31.55*
A^{2}MixStable$(4, 4)$	27.04	33.14*	29.85	17.91	16.21	17.86	31.42*
Jarque-Bera							
MixNormal$(2, 2)$	28.31***	51.49***	18.88***	**2.11**	9.73**	9.71**	80.46***
MixNormal$(3, 2)$	17.25***	3.01	5.69	3.19	14.94***	17.53***	7.01*
MixNormal$(3, 3)$	19.61***	198.32***	52.25***	4.95	22.28***	16.10***	65.26***
MixNormal$(4, 4)$	4.61	70.44***	5.57	5.42	13.52***	20.68***	56.53***
MixGED$(2, 2)$	2.24	9.63**	9.63**	3.14	9.82**	6.64*	19.23***
MixGED$(3, 2)$	15.18***	9.35**	9.94**	3.43	14.12***	3.79	7.23*
MixGED$(3, 3)$	17.54***	12.13***	4.45	4.22	19.36***	5.50	18.01***
MixGED$(4, 4)$	4.97	9.10**	9.06**	4.69	15.94***	6.67*	16.18***
A^{1}MixStable$(2, 2)$	2.06	1.28	8.40**	2.23	10.47**	5.49	0.80
A^{1}MixStable$(3, 2)$	2.04	1.63	2.72	3.31	14.52***	4.14	1.28
A^{1}MixStable$(3, 3)$	1.31	2.47	2.22	5.74	18.98***	4.88	1.71
A^{1}MixStable$(4, 4)$	1.17	0.79	1.45	3.51	10.79***	7.91**	2.23
A^{2}MixStable$(2, 2)$	**0.28**	0.06*	0.54	2.75	3.70	8.56**	0.38
A^{2}MixStable$(3, 2)$	0.32	**0.05****	**0.39**	2.83	3.41	6.10*	1.77
A^{2}MixStable$(3, 3)$	0.67	0.23	0.42	3.31	3.78	**3.13**	**0.03****
A^{2}MixStable$(4, 4)$	0.69	0.25	0.51	3.11	**3.41**	3.25	0.17
Shapiro-Wilk (test statistic ν is transformed by $1000(1 - \nu)$)							
MixNormal$(2, 2)$	3.16***	3.87***	2.65***	**1.23**	3.11**	2.01	6.04***
MixNormal$(3, 2)$	2.32***	0.94	1.68**	1.32*	4.00***	2.91**	1.63**
MixNormal$(3, 3)$	2.40***	6.94***	4.53***	1.65**	4.72***	2.69*	5.37***
MixNormal$(4, 4)$	1.16	4.17***	1.44**	1.60**	3.37**	3.13**	5.43***
MixGED$(2, 2)$	0.92	1.78**	2.23***	1.25	3.10**	1.74	3.05***
MixGED$(3, 2)$	2.27***	1.72**	2.30***	1.27*	3.87***	**1.01**	2.22***
MixGED$(3, 3)$	2.25***	1.93***	1.30*	1.52**	4.49***	1.49	3.68***
MixGED$(4, 4)$	1.19	1.50**	2.24***	1.50**	3.87***	1.55	3.62***
A^{1}MixStable$(2, 2)$	0.67	0.58	1.78**	1.31*	3.25**	1.32	0.91
A^{1}MixStable$(3, 2)$	0.57	0.55	1.00	1.27*	3.93***	1.17	**0.82**
A^{1}MixStable$(3, 3)$	0.57	0.59	0.90	1.56**	4.43***	1.35	1.23
A^{1}MixStable$(4, 4)$	**0.56**	**0.41**	0.81	1.34*	3.12**	1.86	1.43**
A^{2}MixStable$(2, 2)$	1.51**	1.16	0.75	1.26*	**2.06**	2.19	1.76**
A^{2}MixStable$(3, 2)$	1.55**	1.15	**0.74**	1.33*	2.14	1.85	1.73**
A^{2}MixStable$(3, 3)$	1.43**	1.15	0.77	1.27*	2.11	1.34	1.75**
A^{2}MixStable$(4, 4)$	1.44**	1.08	0.76	1.28*	2.14	1.36	1.85***

Table 7: Ljung-Box, Jarque-Bera and Shapiro-Wilk test statistics for all models (with $\delta = 1$) and data sets under study. Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
MixNormal(2, 2)	1.01	1.11	1.21	0.87	1.60**	1.15	**0.96**
MixNormal(3, 2)	0.96	1.08	1.06	0.87	1.68**	1.11	0.60
MixNormal(3, 3)	0.84	1.09	1.13	0.84	1.63**	1.20	0.57
MixNormal(4, 4)	0.92	1.12	1.03	0.89	1.62**	1.19	0.61
MixGED(2, 2)	**1.00**	1.11	1.11	0.90	1.75**	1.40	0.81
MixGED(3, 2)	0.96	1.18	1.07	0.87	1.67**	1.11	0.63
MixGED(3, 3)	0.88	1.10	1.03	0.80	1.76**	1.32	0.59
MixGED(4, 4)	0.88	1.10	1.03	0.85	1.72**	1.20	0.64
A^1MixStable(2, 2)	1.11	1.13	1.11	0.87	1.75**	1.16	0.68
A^1MixStable(3, 2)	1.09	1.17	1.03	0.88	1.60**	1.13	0.62
A^1MixStable(3, 3)	1.02	1.14	1.17	0.85	1.63**	1.22	0.70
A^1MixStable(4, 4)	0.88	1.06	**1.02**	0.92	1.63**	1.20	0.67
A^2MixStable(2, 2)	0.77	**0.98**	1.05	0.90	1.46	1.13	0.71
A^2MixStable(3, 2)	0.77	0.98	1.07	0.91	**1.40**	1.09	0.68
A^2MixStable(3, 3)	0.77	0.98	1.05	**0.98**	1.42	**1.08**	0.64
A^2MixStable(4, 4)	0.78	0.98	1.04	0.93	1.46	1.09	0.66
5% VaR							
MixNormal(2, 2)	5.42	5.48	5.46	5.83**	5.02	4.28	5.49
MixNormal(3, 2)	5.38	5.35	5.22	5.70*	5.15	3.96	4.81
MixNormal(3, 3)	5.23	5.20	5.22	5.58	5.17	4.01	4.60
MixNormal(4, 4)	5.33	**5.04**	5.19	5.56	5.18	3.96	4.45
MixGED(2, 2)	5.30	5.20	5.45	5.90**	**5.02**	4.51	5.09
MixGED(3, 2)	**5.22**	5.34	5.30	5.77**	5.33	4.15	4.61
MixGED(3, 3)	5.26	5.13	5.35	5.80**	5.23	4.44	4.58
MixGED(4, 4)	5.33	5.22	5.30	5.52	5.20	4.24	4.49
A^1MixStable(2, 2)	5.46	5.33	5.15	5.86**	5.10	4.06	**5.04**
A^1MixStable(3, 2)	5.27	5.21	5.05	5.49	5.15	3.95	4.83
A^1MixStable(3, 3)	5.31	5.14	5.05	5.63*	5.15	4.20	4.37
A^1MixStable(4, 4)	5.29	4.94	**5.03**	**5.48**	5.20	4.19	4.26
A^2MixStable(2, 2)	5.47	5.47	5.36	5.67*	5.55	**4.61**	5.61*
A^2MixStable(3, 2)	5.46	5.46	5.32	5.67*	5.55	4.61	5.40
A^2MixStable(3, 3)	5.51	5.40	5.28	5.60*	5.65	4.54	5.37
A^2MixStable(4, 4)	5.51	5.43	5.26	5.64*	5.64	4.53	5.52
10% VaR							
MixNormal(2, 2)	10.69	11.21**	10.60	10.62	9.68	9.07	11.17**
MixNormal(3, 2)	10.75	10.99*	10.97*	10.55	9.32	8.88	10.24
MixNormal(3, 3)	**10.56**	10.71	10.84*	10.64	9.10	8.72	10.16
MixNormal(4, 4)	10.76	**10.63**	10.85*	10.47	9.07	8.79	9.69
MixGED(2, 2)	10.98*	11.00**	10.63	10.61	9.57	8.98	**10.05**
MixGED(3, 2)	10.87*	10.94*	10.90*	10.63	9.36	8.82	9.74
MixGED(3, 3)	10.63	10.76	10.99*	10.53	9.10	8.88	9.70
MixGED(4, 4)	10.73	10.72	10.79	10.53	9.12	8.88	9.67
A^1MixStable(2, 2)	10.80	10.83*	**10.40**	10.53	9.55	9.07	10.42
A^1MixStable(3, 2)	10.68	10.88*	10.85*	**10.24**	9.40	9.04	10.25
A^1MixStable(3, 3)	10.66	10.74	10.68	10.42	9.10	8.59	10.20
A^1MixStable(4, 4)	10.60	10.64	10.69	10.37	9.21	8.59	9.71
A^2MixStable(2, 2)	11.59***	11.26**	10.73	10.99*	9.61	9.15	11.26**
A^2MixStable(3, 2)	11.69***	11.36**	10.77	10.95*	9.61	9.29	11.22**
A^2MixStable(3, 3)	11.74***	11.43***	10.71	10.95*	9.71	**9.29**	11.28**
A^2MixStable(4, 4)	11.77***	11.36**	10.66	10.96*	**9.73**	9.29	11.34**

Table 8: Predicted VaR coverage percentages (point estimates) at the 1%, 5% and 10% level for all models under study (with $\delta = 1$). Entries in boldface denote the best (closest to the true value) estimate. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	$/€	NASDAQ
1% VaR							
MixNormal(2, 2)	0.10	0.22	0.12	0.11	0.32	0.10	**0.06**
MixNormal(3, 2)	0.09	0.15	0.13	0.08	0.34	0.12	0.29
MixNormal(3, 3)	0.08	0.10	0.10	0.10	0.33	0.16	0.31
MixNormal(4, 4)	0.05	0.12	0.08	0.08	0.29	0.17	0.33
MixGED(2, 2)	0.07	0.18	0.11	0.08	0.30	0.22	0.20
MixGED(3, 2)	0.08	0.15	0.12	0.09	0.34	0.10	0.34
MixGED(3, 3)	0.07	0.11	0.14	0.13	0.32	0.20	0.33
MixGED(4, 4)	0.07	0.12	**0.08**	0.11	0.29	0.13	0.36
A^1MixStable(2, 2)	**0.04**	0.17	0.20	0.12	0.33	0.10	0.21
A^1MixStable(3, 2)	0.05	0.16	0.14	0.09	0.34	0.13	0.26
A^1MixStable(3, 3)	0.06	0.16	0.14	0.10	0.30	0.15	0.25
A^1MixStable(4, 4)	0.07	**0.07**	0.10	0.07	0.28	0.21	0.32
A^2MixStable(2, 2)	0.18	0.09	0.09	0.07	0.11	0.07	0.26
A^2MixStable(3, 2)	0.19	0.08	0.10	0.08	**0.10**	**0.05**	0.27
A^2MixStable(3, 3)	0.19	0.11	0.10	**0.07**	0.11	0.05	0.28
A^2MixStable(4, 4)	0.19	0.08	0.09	0.07	0.11	0.05	0.27
5% VaR							
MixNormal(2, 2)	0.19	0.36	0.22	0.31	**0.34**	0.53	**0.20**
MixNormal(3, 2)	0.17	0.25	0.16	0.28	0.47	0.69	0.49
MixNormal(3, 3)	**0.10**	0.23	0.10	0.28	0.47	0.62	0.53
MixNormal(4, 4)	0.20	0.22	0.10	0.27	0.42	0.59	0.64
MixGED(2, 2)	0.20	0.27	0.16	0.31	0.40	**0.31**	0.35
MixGED(3, 2)	0.19	0.29	0.12	0.26	0.46	0.54	0.60
MixGED(3, 3)	0.16	0.23	0.12	0.28	0.47	0.33	0.61
MixGED(4, 4)	0.17	0.22	0.10	0.27	0.44	0.37	0.64
A^1MixStable(2, 2)	0.25	0.26	0.14	0.31	0.42	0.54	0.28
A^1MixStable(3, 2)	0.16	0.24	0.13	0.29	0.45	0.66	0.40
A^1MixStable(3, 3)	0.16	0.18	0.10	0.30	0.49	0.60	0.56
A^1MixStable(4, 4)	0.17	**0.16**	**0.09**	0.28	0.42	0.63	0.65
A^2MixStable(2, 2)	0.30	0.22	0.17	0.25	0.55	0.34	0.31
A^2MixStable(3, 2)	0.30	0.19	0.18	0.26	0.58	0.41	0.35
A^2MixStable(3, 3)	0.31	0.18	0.21	**0.24**	0.57	0.42	0.33
A^2MixStable(4, 4)	0.31	0.19	0.20	0.24	0.59	0.41	0.31
10% VaR							
MixNormal(2, 2)	0.28	0.53	0.45	0.49	**0.30**	0.65	0.59
MixNormal(3, 2)	0.25	0.43	0.44	0.43	0.42	0.78	0.36
MixNormal(3, 3)	**0.17**	0.35	0.40	0.41	0.48	0.83	0.39
MixNormal(4, 4)	0.23	0.29	0.36	0.39	0.45	0.82	0.52
MixGED(2, 2)	0.28	0.44	0.41	0.52	0.36	0.53	0.27
MixGED(3, 2)	0.26	0.50	0.38	0.46	0.42	0.70	0.49
MixGED(3, 3)	0.22	0.36	0.42	0.43	0.48	0.55	0.50
MixGED(4, 4)	0.21	0.37	0.40	0.40	0.45	0.60	0.52
A^1MixStable(2, 2)	0.31	0.33	**0.27**	0.51	0.38	0.66	**0.25**
A^1MixStable(3, 2)	0.21	0.36	0.35	**0.38**	0.41	0.68	0.30
A^1MixStable(3, 3)	0.20	0.28	0.32	0.43	0.46	0.84	0.44
A^1MixStable(4, 4)	0.21	**0.23**	0.30	0.42	0.44	0.86	0.52
A^2MixStable(2, 2)	0.69	0.49	0.46	0.52	0.46	**0.47**	0.73
A^2MixStable(3, 2)	0.68	0.48	0.46	0.51	0.47	0.48	0.71
A^2MixStable(3, 3)	0.72	0.43	0.45	0.51	0.48	0.47	0.70
A^2MixStable(4, 4)	0.73	0.47	0.45	0.52	0.49	0.48	0.76

Table 9: Integrated root mean squared error of the VaR prediction up to the 1%, 5% and 10% level for all models under study (with $\delta = 1$). Entries in boldface denote the best estimate. Results are based on the same out-of-sample exercise as in Table 3.

model	DAX	S&P	DJIA	NIKKEI	¥/€	\$/€	NASDAQ
Unconditional Coverage, LR_{UC}							
MixNormal(2, 2)	**3.1e-4**	0.14	0.88	0.68	3.23*	0.28	**0.05**
MixNormal(3, 2)	0.05	0.14	0.03	0.68	4.14**	**0.07**	5.62**
MixNormal(3, 3)	1.07	0.14	0.32	1.07	4.14**	0.28	6.81***
MixNormal(4, 4)	0.39	0.32	0.03	0.39	4.14**	0.28	5.62**
MixGED(2, 2)	**3.1e-4**	0.14	0.32	0.39	5.16**	1.70	1.07
MixGED(3, 2)	0.05	0.56	0.14	0.68	4.14**	**0.07**	4.57**
MixGED(3, 3)	0.68	0.14	**3.1e-4**	1.56	5.16**	1.11	5.62**
MixGED(4, 4)	0.68	0.14	**3.1e-4**	0.68	5.16**	0.28	4.57**
A^1MixStable(2, 2)	0.32	0.32	0.14	0.68	5.16**	0.28	3.65*
A^1MixStable(3, 2)	0.14	0.56	**3.1e-4**	0.68	3.23*	**0.07**	4.57**
A^1MixStable(3, 3)	**3.1e-4**	0.32	0.56	0.68	4.14**	0.28	2.84*
A^1MixStable(4, 4)	0.68	**0.03**	**3.1e-4**	0.39	4.14**	0.28	3.65*
A^2MixStable(2, 2)	1.56	0.05	0.03	0.39	**1.70**	**0.07**	2.84*
A^2MixStable(3, 2)	1.56	0.05	0.14	0.39	**1.70**	**0.07**	3.65*
A^2MixStable(3, 3)	1.56	0.05	0.03	**0.05**	**1.70**	**0.07**	4.57**
A^2MixStable(4, 4)	1.56	0.05	0.03	0.17	**1.70**	**0.07**	3.65*
Independence, LR_{IND}							
MixNormal(2, 2)	1.28	0.63	0.77	0.39	1.06	0.37	0.50
MixNormal(3, 2)	1.40	0.63	0.59	0.39	0.92	**0.33**	0.19
MixNormal(3, 3)	1.98	0.63	0.67	0.36	0.92	0.37	**0.16**
MixNormal(4, 4)	1.67	0.67	0.59	0.43	0.92	0.37	0.19
MixGED(2, 2)	1.28	0.63	0.67	0.43	0.79	0.53	0.36
MixGED(3, 2)	1.40	0.72	0.63	0.39	0.92	**0.33**	0.21
MixGED(3, 3)	1.82	0.63	**0.54**	**0.32**	0.79	0.48	0.19
MixGED(4, 4)	1.82	0.63	**0.54**	0.39	0.79	0.37	0.21
A^1MixStable(2, 2)	**0.96**	0.67	0.63	0.39	0.79	0.37	0.24
A^1MixStable(3, 2)	1.06	0.72	**0.54**	0.39	1.06	**0.33**	0.21
A^1MixStable(3, 3)	1.28	0.67	0.72	0.39	0.92	0.37	0.26
A^1MixStable(4, 4)	1.82	0.59	**0.54**	0.43	0.92	0.37	0.24
A^2MixStable(2, 2)	2.15	**0.50**	0.59	0.43	1.37	**0.33**	0.26
A^2MixStable(3, 2)	2.15	**0.50**	0.63	0.43	**0.53**	**0.33**	0.24
A^2MixStable(3, 3)	2.15	**0.50**	0.59	0.50	1.37	**0.33**	0.21
A^2MixStable(4, 4)	2.15	**0.50**	0.59	0.46	1.37	**0.33**	0.24
Conditional Coverage, LR_{CC}							
MixNormal(2, 2)	1.28	0.77	1.65	1.08	4.28	0.66	**0.55**
MixNormal(3, 2)	1.45	0.77	0.62	1.08	5.06	**0.40**	5.81
MixNormal(3, 3)	3.05	0.77	0.99	1.43	5.06	0.66	6.97*
MixNormal(4, 4)	2.06	0.99	0.62	0.81	5.06	0.66	5.81
MixGED(2, 2)	1.28	0.77	0.99	0.81	5.95	2.24	1.43
MixGED(3, 2)	1.45	1.29	0.77	1.08	5.06	**0.40**	4.78
MixGED(3, 3)	2.50	0.77	**0.54**	1.89	5.95	1.58	5.81
MixGED(4, 4)	2.50	0.77	**0.54**	1.08	5.95	0.66	4.78
A^1MixStable(2, 2)	1.28	0.99	0.77	1.08	5.95	0.66	3.88
A^1MixStable(3, 2)	**1.20**	1.29	**0.54**	1.08	4.28	**0.40**	4.78
A^1MixStable(3, 3)	1.28	0.99	1.29	1.08	5.06	0.66	3.11
A^1MixStable(4, 4)	2.50	0.62	**0.54**	0.81	5.06	0.66	3.88
A^2MixStable(2, 2)	3.71	**0.55**	0.62	0.81	3.08	**0.40**	3.11
A^2MixStable(3, 2)	3.71	**0.55**	0.77	0.81	**2.24**	**0.40**	3.88
A^2MixStable(3, 3)	3.71	**0.55**	0.62	**0.55**	3.08	**0.40**	4.78
A^2MixStable(4, 4)	3.71	**0.55**	0.62	0.64	3.08	**0.40**	3.88

Table 10: Test statistics at the 1%-VaR level, $LR_{CC} = LR_{UC} + LR_{IND}$, as described in Christoffersen (1998) for all models under study (with $\delta = 1$). Entries in boldface denote the best outcomes. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Results are based on the same out-of-sample exercise as in Table 3.